

**(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)**

**(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international**



**(43) Date de la publication internationale
2 mai 2002 (02.05.2002)**

PCT

**(10) Numéro de publication internationale
WO 02/34903 A2**

(51) Classification internationale des brevets⁷ :
C12N 15/00, C07K 14/705

(21) Numéro de la demande internationale :
PCT/FR01/03219

(22) Date de dépôt international :
17 octobre 2001 (17.10.2001)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
00/13649 24 octobre 2000 (24.10.2000) FR
60/253,141 28 novembre 2000 (28.11.2000) US

(71) Déposants (pour tous les États désignés sauf US) : AVEN-TIS PHARMA S.A. [FR/FR]; 20 avenue Raymond Aron, F-92160 Antony (FR). INSERM [FR/FR]; 101 rue de Tolbiac, F-75654 Paris Cedex 13 (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : DENEFLE, Patrice [FR/FR]; 45 avenue des Fusillés de Chateaubriand, F-94100 Saint Maur (FR). ROSIER, Marie-Françoise [FR/FR]; 21 rue des Baconnets, F-92160 Antony (FR).

(81) États désignés (national) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) États désignés (régional) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Suite sur la page suivante]

(54) Title: NUCLEIC ACID GENERATING THE ABCA7 GENE, MOLECULES MODULATING ITS ACTIVITY AND THERAPEUTIC APPLICATIONS

(54) Titre : ACIDE NUCLEIQUE REGULATEUR DU GENE ABCA7, MOLECULES MODULANT SON ACTIVITE ET APPLICATIONS THERAPEUTIQUES

(57) Abstract: The invention concerns a nucleic acid capable of regulating the ABCA7 gene transcription, which codes for a carrier protein capable of intervening in the metabolism of lipids and/or in the process involving the immune system and inflammation. In addition, owing to the position of the gene on the chromosome 19 in q 13, ABCA7 is potentially involved in other pathologies genetically related to said locus. The invention also concerns nucleotide constructs comprising a polynucleotide coding for a polypeptide or a nucleic acid of interest, placed under the control of a nucleic acid regulating the ABCA7 gene. The invention further concerns recombinant vectors, transformed host cells and non-human transgenic mammals comprising a nucleic acid regulating the ABCA7 gene transcription and or said nucleotide construct, and methods for screening molecules or substances capable of modulating the activity of the nucleic acid regulating the ABCA7 gene.

(57) Abrégé : La présente invention concerne un acide nucléique capable de réguler la transcription de gène ABCA7, qui code pour une protéine transporteur susceptible de jouer un rôle dans le métabolisme des lipides et/ou dans les processus impliquant le système immunitaire et l'inflammation. Par ailleurs, de par la position du gène sur le chromosome 19 en q13, ABCA7 est potentiellement impliqué dans d'autres pathologies liées génétiquement à ce locus. La présente invention est également relative à des constructions nucléotidiques comprenant un polynucléotide codant pour un polypeptide ou un acide nucléique d'intérêt, placé sous le contrôle d'un acide nucléique régulateur du gène ABCA7. L'invention a également trait à des vecteurs recombinants, des cellules hôtes transformées et des mammifères transgéniques non humains comprenant un acide nucléique régulateur de la transcription du gène ABCA7 ou une construction nucléotidique précitée, ainsi que des procédés pour le criblage de molécules ou de substances capables de moduler l'activité de l'acide nucléique régulateur de gène ABCA7.

WO 02/34903 A2



Publiée :

— *sans rapport de recherche internationale, sera republiée dès réception de ce rapport*

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

ACIDE NUCLEIQUE REGULATEUR DU GENE ABCA7, MOLECULES MODULANT SON ACTIVITE ET APPLICATIONS THERAPEUTIQUES

La présente invention concerne un acide nucléique capable de réguler la transcription du gène ABCA7, qui est un gène susceptible d'intervenir dans le métabolisme des lipides au niveau des tissus hématopoïétiques, ainsi que dans les mécanismes de signalisation cellulaire liés à la réaction immunitaire et à l'inflammation.

La présente invention décrit aussi des polypeptides et des polynucléotides dont une altération de la séquence ou de l'expression est potentiellement impliquée dans des maladies associées au locus génétique q13 du chromosome 19.

La présente invention est également relative à des constructions nucléotidiques comprenant un polynucléotide codant pour un polypeptide ou produisant un acide nucléique d'intérêt, placé sous le contrôle d'un acide nucléique régulateur du gène humain ou murin ABCA7.

L'invention a également trait à des vecteurs recombinants, des cellules hôtes transformées, et des mammifères transgéniques non humains, comprenant un acide nucléique régulateur de la transcription du gène ABCA7 humain et de souris ou une construction nucléotidique précitée, ainsi que des procédés pour le criblage de molécules ou de substances capables de moduler l'activité de l'acide nucléique régulateur du gène ABCA7.

L'invention est en outre relative à des procédés permettant de détecter une altération de la transcription du gène ABCA7 et ainsi de diagnostiquer un éventuel dysfonctionnement dans le métabolisme lipidique au niveau des tissus hématopoïétiques et dans les mécanismes de signalisation cellulaire de l'immunité.

Elle a également pour objet des substances ou molécules modulant l'activité de l'acide nucléique régulateur de la transcription du gène ABCA7 ainsi que des compositions pharmaceutiques contenant de telles substances ou de telles molécules.

Les protéines transporteurs ABC (ATP-Binding Cassette) constituent une super-famille extrêmement conservées au cours de l'évolution, de la bactérie à l'homme. Ces protéines sont impliquées dans le transport membranaire de divers substrats, par exemple des ions, des acides aminés, des peptides, des sucres, des vitamines ou encore des hormones stéroïdiennes (Higgins et al., *Annu Rev Cell Biol*, 8, (1992) 67-113).

La caractérisation de la séquence complète en acides aminés de certains transporteurs ABC a permis de définir une structure générale commune, comprenant notamment, deux repliements de liaison aux nucléotides (Nucleotide Binding Fold ou NBF) avec des motifs de type Walker A et B, ainsi que deux domaines transmembranaires, chacun des domaines transmembranaires étant constitué de six hélices (Klein et al., *BBA*, 1461 (1999), 237-262). La spécificité des transporteurs ABC pour les différentes molécules transportées semble être déterminée par la structure des domaines transmembranaires, alors que l'énergie nécessaire à l'activité de transport est fournie par la dégradation de l'ATP au niveau du repliement NBF (Dean et al., 10 *Curr.Opin.Genet.Dev.*, 5 (1995) 779-785).

Plusieurs protéines transporteurs ABC ont été identifiées chez l'homme et un certain nombre d'entre elles ont été associées à diverses maladies.

Par exemple, la mucoviscidose est provoquée par des mutations dans le gène CFTR (cystic fibrosis transmembrane conductance regulator), également désigné 15 ABCC7.

Par ailleurs, certains phénotypes de résistance multiple aux médicaments dans les cellules tumorales ont été associés à des mutations dans les gènes codant des protéines MDR (multi-drug resistance) également désignés ABCB, qui ont également une structure de transporteur ABC.

20 D'autres transporteurs ABC ont été associés à des affections neuronales et tumorales (brevet US N°5,858,719) ou sont encore potentiellement impliqués dans des maladies provoquées par une altération de l'homéostasie des métaux, notamment la protéine ABC-3.

De même, un autre ABC transporteur, désigné PFIC2 ou ABCB11, semble 25 être impliqué dans une forme de cholestasie intrahépatique familiale progressive, cette protéine étant potentiellement responsable, chez l'homme, de l'exportation des sels biliaires.

Une sous-famille A des transporteurs ABC désignée ABCA a été également identifiée. Elle se caractérise par la présence d'un segment hautement hydrophobe 30 (HH1 : highly hydrophobic) entre les deux domaines transmembranaires, liés aux deux motifs NBF (Broccardo et al., *BBA* 1461 (1999) 395-404). Quatre membres de cette sous famille ont été jusqu'à présent caractérisés. Il s'agit des transporteurs ABCA1 et ABCA2, tous deux localisés sur le chromosome 9, respectivement aux locus 9q22-9q31

et 9q34, ainsi que le transporteur ABCA3 localisé sur le chromosome 16p13.3, et enfin le transporteur ABCA4 ou ABCR localisé sur le chromosome 1p22 (Broccardo et al., 1999). Les membres de cette sous famille sont également fortement conservés au cours de l'évolution des eucaryotes multicellulaires. A titre d'exemples les transporteurs 5 ABCA1 et ABCA4 qui sont les mieux connus présentent une identité respectivement de 95% et de 88% avec leurs orthologues murins. Les membres de cette sous famille sont en outre fortement apparentés, puisque par exemple les transporteurs ABCA1 et ABCA4 présentent une identité de séquence protéique de 50.9%, ainsi qu'une organisation génomique très similaire (Allikmets et al., *Nat. Genet.* (1997) 15, 236-10 246 ; Broccardo et al., *Biochim. Biophys. Acta* (1999) 1461, 395-404 ; Luciani et al., *Genomics* (1994) 21(1), 150-9 ; Remaley et al., *Proc. Natl. Acad. Sci. USA* (1999) 96(22), 12685-90).

Par ailleurs, les membres de la sous-famille A semblent présenter une spécialisation fonctionnelle similaire au niveau du transport des lipides et des 15 phospholipides membranaires. Il a en effet été montré que la perte de la fonction de ces transporteurs affecte le renouvellement des phospholipides de la bicoche des membranes cellulaires. Dans le cas de ABCA4, on constate dans un premier temps un renouvellement anormal des phosphatidyléthanolamine (PE) dans la partie externe de la membrane des cellules à bâtonnets, qui conduit par une succession d'évènements, à 20 une perte totale de l'acuité visuelle (Weng et al., *Cell* (1999) 98(1), 13-23. Dans le cas de ABCA1, on constate une distribution anormale des phospholipides membranaires au sein des couches de la membrane plasmique, qui résulte plus précisément en la présence en plus grande quantité de phosphatidylsérine dans la couche externe, et en une perturbation de la concentration de Ca^{2+} .

25 Les transporteurs ABCA1 et ABCA4 ont été particulièrement étudiés. Le gène ABCA1 semble en effet être impliqué dans les pathologies liées à un dysfonctionnement du métabolisme du cholestérol induisant des maladies comme l'athérosclérose, ou des déficiences familiales en HDL (FHD), comme la maladie de Tangier (FR 99/7684000 ; Rust et al., *Nat. Genet.*, 22 (1999) 352-355; Brooks-Wilson et 30 al., *Nat. Genet.*, 22 (1999) 336-345; Bodzioch et al., *Nat. Genet.* 22 (1999) 347-351 ; Orso et al., *Nat. Genet.*, 24 (2000) 192-196). La maladie de Tangier semblerait être liée à un déficit cellulaire dans la translocation du cholestérol cellulaire qui entraîne une dégradation des HDLs, et par là même une perturbation du métabolisme lipoprotéique.

Ainsi, il semblerait que les particules HDL qui n'incorporent pas de cholestérol à partir des cellules périphériques, ne sont pas métabolisées correctement, mais sont au contraire éliminées rapidement de l'organisme. La concentration plasmatique en HDL de ces patients est donc extrêmement réduite et les HDLs n'assurent plus le retour du cholestérol vers le foie. Ce cholestérol s'accumule dans ces cellules périphériques et provoque des manifestations cliniques caractéristiques telles que la formation d'amygdales orangées. De plus, d'autres perturbations lipoprotéiques comme une surproduction de triglycérides ainsi qu'une synthèse et un catabolisme intracellulaire accusé des phospholipides sont observées.

Le transporteur ABCA4 a par ailleurs été associé aux maladies dégénératives et inflammatoires oculaires telles que la maladie récessive de Stargardt (Allikmets et al., 1997) et la dégénérescence de la région maculaire de la rétine liée à l'âge (AMD) (Allikmets et al., *Nat.Genet.* **15** (1997) 236-246 ; Allikmets et al., *Science*, **277** (1997) 1805-1807 ; Cremers et al., *Hum. Mol. Genet.* (1998), **7**(3), 355-62 ; Martinez-Mir et al., *Nat. Genet.* **18** (1998) 11-12; Weng et al., *Cell* (1999) **98**(1), 13-23).

Chez l'homme, un ADNc comprenant la totalité de la phase de lecture ouverte d'un nouveau membre de la sous famille A des transporteurs ABC (« ATP-Binding Cassette») a été récemment cloné à partir d'ARN de macrophages humains, et est désigné ABCA7 (Kaminski et al., *BBR*, **273**(2000), 532-538).

La caractérisation de la séquence complète en acides aminés de ABCA7 indique que le produit protéique présente la structure générale caractéristique des transporteurs ABCA, en ce qu'elle comporte la structure symétrique comprenant les deux domaines transmembranaires et deux motifs NBF. En plus de ces motifs caractéristiques, la protéine ABCA7 présente d'autres motifs qui ont été récemment identifiés comme étant caractéristiques des ABCA.

Par ailleurs, la protéine transporteur ABCA7 semble présenter un profil de régulation dépendant des flux de stérol, similaire à celui des autres membres de la sous famille A, et notamment du transporteur ABCA1 (Langman et al., *BBR Com* ; 257(1999), 29-33 ; Laucken et al., *PNAS*, 97(2000) 817-822). Il a été en effet observé 5 par Kaminski et al. (supra) une augmentation de l'expression de ABCA7 après incubation des macrophages humains en présence de lipoprotéines de faible densité acétylées (AcLDL) qui induisent une charge de stérol, ainsi qu'une diminution de l'expression en présence de l'accepteur de cholestérol HDL3 qui provoque une diminution de la charge en stérol.

10 D'autre part, ABCA7 présente comme les autres membres ABCA une certaine spécialisation de son expression tissulaire, le messager de ABCA7 étant présent de manière prédominante dans les tissus hématopoïétiques constitués par les lymphocytes, les granulocytes, le thymus, la rate, la moelle osseuse, ou les tissus foetaux, alors que l'expression de ABCA1 est prédominante dans les macrophages et 15 le placenta, et que celle de ABCA4 est restreinte dans la rétine (Rust et al., *Nat. Genet.*, 22, (1999) 352-355).

20 L'ensemble des données précédemment exposées relatives à l'identité des séquences protéiques, au mécanisme de régulation et la spécificité de l'expression suggère que le gène ABCA7 constitue un autre transporteur de la sous famille A, et que celui-ci présente une fonction similaire ou même redondante à celle des autres transporteurs et notamment à celle du transporteur ABCA1. Ce transporteur interviendrait donc vraisemblablement en tant que médiateur dans le métabolisme des lipides, et il est très possible qu'il soit, au même titre que le transporteur ABCA1, responsable de certains dysfonctionnements ou déficiences métaboliques. Par ailleurs, 25 la spécialisation de l'expression du transporteur ABCA7 indique vraisemblablement que ce dernier joue un rôle dans le transport transmembranaire (export) des lipides dans les tissus hématopoïétiques, et possiblement dans les mécanismes de signalisation lymphocytaires de l'immunité, par exemple dans le cas de la pathogenèse de l'athérosclérose comme l'indiquent Kaminski et al. (Supra).

30 Bien que l'expression du gène ABCA7 humain semble être régulée selon le type de cellule ou encore la situation métabolique d'un type cellulaire donné, la ou les séquences permettant de réguler ce gène n'étaient pas connues.

Or, il existe un besoin dans l'état de la technique d'identifier ces séquences régulatrices pour les raisons suivantes:

5 a) Ces séquences sont susceptibles d'être mutées chez des patients atteints d'une pathologie liée à un déficit dans le transport des lipides, possibles substrats de la protéine ABCA7 ou chez les patients susceptibles de développer de telles pathologies.

La caractérisation des séquences régulatrices du gène ABCA7 humain permettrait de détecter des mutations chez des patients, en particulier de diagnostiquer les individus appartenant à des groupes familiaux à risques. En outre, l'isolement de ces séquences régulatrices rendrait possible la complémentation de la séquence mutée 10 par une séquence fonctionnelle susceptible de pallier les dysfonctionnements métaboliques induits par la ou les mutations diagnostiquées, grâce à la construction de moyens thérapeutiques ciblés, tels que des moyens destinés à la thérapie génique.

15 b) La caractérisation des séquences régulatrices du gène ABCA7 mettrait à la disposition de l'homme du métier des moyens aptes à permettre la construction par génie génétique puis l'expression de gènes déterminés dans les types cellulaires dans lesquels le gène ABCA7 est préférentiellement exprimé.

20 c) Par ailleurs, certaines parties des séquences régulatrices du gène ABCA7 pourraient constituer des séquences promotrices constitutives à fort niveau d'expression, de nature à permettre la construction de moyens nouveaux pour l'expression de séquences déterminées dans les cellules, complétant un ensemble de moyens déjà existants.

Force est de constater que, malgré les efforts entrepris, les séquences régulatrices du gène ABCA7 étaient restées jusqu'à ce jour totalement inconnues.

25 Les inventeurs ont désormais isolé puis analysé un ADN génomique humain de 33,5kb comprenant les 46 exons du cadre ouvert de lecture du gène ABCA7 ainsi que la région non transcrive d'environ 1,1kb localisée du côté 5' de l'exon 1, en amont du site +1 de transcription, et comprenant des signaux de régulation du gène ABCA7 humain.

30 Les inventeurs ont également isolé puis analysé un ADN génomique murin de 20Kb comprenant les 45 exons du cadre ouvert de lecture du gène ABCA7 ainsi que la région non transcrive d'environ 1,2Kb chez la souris localisée du côté 5' de l'exon 1, en amont du site +1 de transcription, et comprenant des signaux de régulation du gène ABCA7 murin.

DEFINITIONS GENERALES

5 Le terme "isolé" au sens de la présente invention désigne un matériel biologique (acide nucléique ou protéine) qui a été soustrait à son environnement original (l'environnement dans lequel il est localisé naturellement).

Par exemple un polynucléotide présent à l'état naturel dans une plante ou un animal n'est pas isolé. Le même polynucléotide séparé des acides nucléiques 10 adjacents au sein desquels il est naturellement inséré dans le génome de la plante ou l'animal est considéré comme "isolé".

Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeurer néanmoins à l'état isolé du fait que le vecteur ou la composition ne constitue pas son environnement 15 naturel.

Le terme "purifié" ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusive de la présence d'autres composés. Il s'agit plutôt d'une définition relative.

Un polynucléotide est à l'état "purifié" après purification du matériel de départ 20 ou du matériel naturel d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

Aux fins de la présente description, l'expression "séquence nucléotidique" peut être employée pour désigner indifféremment un polynucléotide ou un acide nucléique. L'expression "séquence nucléotidique" englobe le matériel de génétique lui-même et n'est donc pas restreinte à l'information concernant sa séquence. 25

Les termes "acide nucléique", "polynucléotide", "oligonucléotide" ou encore "séquence nucléotidique" englobent des séquences d'ARN, d'ADN, d'ADNc ou encore des séquences hybrides ARN/ADN de plus d'un nucléotide, indifféremment sous la forme simple chaîne ou sous la forme de duplex.

30 Le terme "nucléotide" désigne à la fois les nucléotides naturels (A, T, G, C) ainsi que des nucléotides modifiés qui comprennent au moins une modification telle que (1) un analogue d'une purine, (2) un analogue d'une pyrimidine, ou (3) un sucre

analogue, des exemples de tels nucléotides modifiés étant décrits par exemple dans la demande PCT N°WO 95/04 064.

Aux fins de la présente invention, un premier polynucléotide est considéré comme étant "complémentaire" d'un second polynucléotide lorsque chaque base du 5 premier nucléotide est appariée à la base complémentaire du second polynucléotide dont l'orientation est inversée. Les bases complémentaires sont A et T (ou A et U), ou C et G.

Par "variant" d'un acide nucléique selon l'invention, on entendra un acide nucléique qui diffère d'une ou plusieurs bases par rapport au polynucléotide de 10 référence. Un acide nucléique variant peut être d'origine naturelle, tel qu'un variant allélique retrouvé naturellement, ou peut être aussi un variant non naturel obtenu par exemple par des techniques de mutagenèse.

En général, les différences entre l'acide nucléique de référence et l'acide nucléique variant sont réduites de telle sorte que les séquences nucléotidiques de 15 l'acide nucléique de référence et de l'acide nucléique variant sont très proches et, dans de nombreuses régions, identiques. Les modifications de nucléotides présentes dans un acide nucléique variant peuvent être silencieuses, ce qui signifie qu'elles n'altèrent pas les séquences d'aminoacides codées par ledit acide nucléique variant.

Cependant, les changements de nucléotides dans un acide nucléique variant 20 peuvent aussi produire des substitutions, additions, délétions dans le polypeptide codé par l'acide nucléique variant par rapport aux peptides codés par l'acide nucléique de référence. En outre, de telles modifications de nucléotides dans les régions codantes peuvent produire des substitutions, conservatives ou non conservatives dans la séquence d'aminoacides.

25 De préférence, les acides nucléiques variants selon l'invention codent pour des polypeptides qui conservent sensiblement la même fonction ou activité biologique que le polypeptide de l'acide nucléique de référence ou encore la capacité à être reconnus par des anticorps dirigés contre les polypeptides codés par l'acide nucléique initial.

30 Certains acides nucléiques variants coderont ainsi pour des formes mutées des polypeptides dont l'étude systématique permettra de déduire des relations structure activité des protéines en question. La connaissance de ces variants par rapport à la

maladie étudiée est fondamentale puisqu'elle permet de comprendre la cause moléculaire de la pathologie.

On entendra par "fragment" un acide nucléique de référence selon l'invention, une séquence nucléotidique de longueur réduite par rapport à l'acide nucléique de référence et comprenant, sur la partie commune, une séquence en nucléotides identique à l'acide nucléique de référence.

Un tel "fragment" d'acide nucléique selon l'invention peut être le cas échéant, compris dans un polynucléotide plus grand duquel il est constitutif.

De tels fragments comprennent, ou alternativement consistent en, des oligonucléotides de longueur allant de 20 à 25, 30, 40, 50, 70, 80, 100, 200, 500, 1000 ou 1500 nucléotides consécutifs d'un acide nucléique selon l'invention.

Par "fragment biologiquement actif" d'un acide régulateur de la transcription selon l'invention, on entend un acide nucléique capable de moduler la transcription d'une séquence d'ADN placée sous son contrôle. Un tel fragment biologiquement actif comprend un promoteur de base et/ou un élément régulateur, tels que définis dans la présente description.

Par "acide nucléique régulateur" selon l'invention, on entend un acide nucléique qui active et/ou régule l'expression d'une séquence d'ADN sélectionnée et placée sous son contrôle.

Par "promoteur", on entend une séquence d'ADN reconnue par les protéines de la cellule impliquées dans l'initiation de la transcription d'un gène. Le promoteur de base est l'acide nucléique régulateur minimal capable d'initier la transcription d'une séquence d'ADN déterminée qui est placée sous son contrôle. En général, le promoteur de base consiste en une région d'ADN génomique en amont du site d'initiation de la transcription où se trouve très souvent une séquence CAAT (où se fixe un ou plusieurs facteurs protéiques de transcription) ainsi que, sauf dans de rares cas comme dans certains gènes domestiques, la séquence TATA ou "TATA box" ou une boîte apparentée. C'est au niveau de cette boîte que se fixe une ARN polymérase ainsi qu'un ou plusieurs facteurs de transcription, tels que les protéines se fixant sur la boîte "TATA" (TATA box Binding Proteins ou TBPs).

Une séquence nucléotidique est "placée sous le contrôle" d'un acide nucléique régulateur lorsque cet acide nucléique régulateur est localisé, par rapport à la

séquence nucléotidique, de telle manière à contrôler l'initiation de la transcription de la séquence nucléotidique par une ARN polymérase.

Par "élément régulateur" ou "séquence régulatrice" au sens de l'invention, on entend un acide nucléique comprenant des éléments capables de moduler la transcription initiée par un promoteur de base, tels que des sites de fixation de divers facteurs de transcription, des séquences "enhancer" d'augmentation de la transcription ou des séquences "silencer" d'inhibition de la transcription.

Par séquence "enhancer", on entend une séquence d'ADN incluse dans un acide nucléique régulateur capable d'augmenter ou de stimuler la transcription initiée par un promoteur de base.

Par séquence "silencer", on entend une séquence d'ADN incluse dans un acide régulateur capable de diminuer ou d'inhiber la transcription initiée par un promoteur de base.

Des éléments régulateurs peuvent être présents en dehors de la séquence localisée du côté 5' du site d'initiation de la transcription, par exemple dans les introns et les exons, y compris dans les séquences codantes.

Le promoteur de base et l'élément régulateur peuvent être "spécifiques d'un ou plusieurs tissus", s'ils permettent la transcription d'une séquence d'ADN déterminée, placée sous leur contrôle, préférentiellement dans certaines cellules (par exemple les cellules spécifiques d'un tissu), c'est à dire soit exclusivement dans les cellules de certains tissus, soit à des niveaux de transcription différents selon les tissus.

Par "facteur de transcription", on entend des protéines qui interagissent préférentiellement avec des éléments régulateurs d'un acide nucléique régulateur selon l'invention, et qui stimulent ou au contraire répriment la transcription. Certains facteurs de transcription sont actifs sous forme de monomères, d'autres étant actifs sous la forme d'homo- ou d'hétérodimères.

Le terme "modulation" vise soit une régulation positive (augmentation, stimulation) de la transcription, soit une régulation négative (diminution, inhibition, blocage) de la transcription.

Le "pourcentage d'identité" entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptide dans la fenêtre de comparaison peut ainsi comprendre des additions ou des délétions (par exemple des "gaps") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

5 Le pourcentage est calculé en déterminant le nombre de positions auquel une base nucléique ou un résidu d'aminoacide identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auquel il y a identité entre les deux bases ou résidus d'aminoacides par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat 10 par 100 afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus contenus dans le package de la Société WISCONSIN GENETICS SOFTWARE PACKAGE, GENETICS COMPUTER GROUP (GCG), 575 Science Doctor, Madison, WISCONSIN.

15 À titre d'illustration, le pourcentage d'identité de séquence pourra être effectué à l'aide du logiciel BLAST (versions BLAST 1.4.9 de mars 1996, BLAST 2.0.4 de février 1998 et BLAST 2.0.6 de septembre 1998), en utilisant exclusivement les paramètres par défaut (Altschul et al, *J. Mol. Biol.*, (1990) 215 : 403-410 ; Altschul et al, *Nucleic Acids Res.* (1997) 25 : 3389-3402). Blast recherche des séquences 20 similaires/homologues à une séquence "requête" de référence, à l'aide de l'algorithme d'Altschul et al. (Supra). La séquence requête et les bases de données utilisées peuvent être peptidiques ou nucléiques, toute combinaison étant possible.

Par "conditions d'hybridation de forte stringence" au sens de la présente invention, on entendra les conditions suivantes :

25

1- Compétition des membranes et PRE HYBRIDATION :

- Mélanger : 40µl ADN sperme de saumon (10mg/ml)

+ 40 µl ADN placentaire humain (10mg/ml)

30

- Dénaturer 5 mn à 96°C, puis plonger dans la glace le mélange.

- Oter le tampon SSC 2X et verser 4 ml de mix formamide dans le tube d'hybridation contenant les membranes.

- Ajouter le mélange des deux ADNs dénaturés.

5

- Incubation à 42°C pendant 5 à 6 heures, avec rotation.

2- Compétition de la sonde marquée :

10 - Ajouter à la sonde marquée et purifiée 10 à 50 µl ADN Cot I, selon la quantité d'hybridations non spécifiques.

- Dénaturer 7 à 10 mn à 95°C.

15 - Incuber à 65°C pendant 2 à 5 heures.

3- Hybridation:

- Oter le mix de pré hybridation.

20

- Mélanger 40 µl ADN sperme de saumon + 40 µl ADN placentaire humain ; dénaturer 5 mn à 96°C, puis plonger dans la glace.

25 - Ajouter dans le tube d'hybridation 4 ml de mix formamide, le mélange des deux ADN et la sonde marquée/ADN Cot I dénaturée.

- Incuber 15 à 20 heures à 42°C, avec rotation.

4- Lavages :

30

- Un lavage à température ambiante dans du SSC 2X, pour rincer.

- 2 fois 5 minutes à température ambiante SSC 2X et SDS 0,1%.

- 2 fois 15 minutes SSC 0,1X et SDS 0,1% à 65°C.

Envelopper les membranes dans du Saran et exposer.

Les conditions d'hybridation décrites plus haut sont adaptées à l'hybridation dans des conditions de forte stringence, d'une molécule d'acide nucléique d'une longueur variable de 20 nucléotides à plusieurs centaines de nucléotides.

Il va sans dire que les conditions d'hybridation ci-dessus décrites peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon des techniques connues de l'homme du métier.

Les conditions convenables d'hybridation peuvent par exemple être adaptées selon l'enseignement contenu dans l'ouvrage de HAMES et HIGGINS (1985) (Nucleic acid Hybridization iapractical Approach, Hames and Higgins Ed., IRL Press, Oxford) ou encore dans l'ouvrage de F. AUSUBEL et al (1999) (Currents Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.).

Par " transformation " au sens de l'invention, on entend l'introduction d'un acide nucléique (ou d'un vecteur recombinant) dans une cellule hôte. Le terme " transformation " englobe également une situation dans laquelle le génotype d'une cellule a été modifié par un acide nucléique exogène, et que cette cellule ainsi transformée exprime ledit acide nucléique exogène, par exemple sous la forme d'un polypeptide recombinant ou encore sous la forme d'un acide nucléique sens ou antisens.

Par " animal transgénique " au sens de l'invention, on entend un animal non humain, de préférence un mammifère, dans lequel une ou plusieurs cellules contiennent un acide nucléique hétérologue introduit grâce à l'intervention humaine, tel que par des techniques de transgénèse bien connues de l'homme du métier. L'acide nucléique hétérologue est introduit directement ou indirectement dans la cellule ou le précurseur de la cellule, par manipulation génétique telle que micro-injection ou infection par un virus recombinant. L'acide nucléique hétérologue peut être intégré dans le chromosome, ou peut se présenter sous la forme d'ADN se répliquant de manière extra-chromosomique.

ACIDE NUCLEIQUE REGULATEUR DU GENE ABCA7

A partir de banques de vecteurs de type BAC préparées à partir de matériel génomique humain et murin, les inventeurs ont réussi à isoler un acide nucléique régulateur des gènes ABCA7 humain et murin.

Les inventeurs ont déterminé par une analyse comparative des séquences 5 génomiques humaines et murines, un acide nucléique régulateur comprenant particulièrement deux modules de régulation conservés chez l'homme et la souris. Les inventeurs ont donc déterminé que l'acide nucléique régulateur de la transcription du gène ABCA7, lorsqu'il est défini de la manière la plus large, est constitué d'un polynucléotide comprenant, de l'extrémité 5' vers l'extrémité 3':

10 - une région non transcrrite d'environ 1,2 kb localisée en amont du site d'initiation de la transcription du gène ABCA7, et
- la séquence partielle du premier exon du gène ABCA7.

Dans sa définition la plus générale, l'acide nucléique régulateur de la transcription du gène ABCA7 comprend l'ensemble des régions nucléotidiques telles 15 que définies ci-dessus et est identifié dans la séquence SEQ ID N°1 selon l'invention.

Ainsi, un premier objet de l'invention consiste en un acide nucléique comprenant un polynucléotide ayant au moins 20 nucléotides consécutifs de la séquence nucléotidique SEQ ID N°1, ou un acide nucléique de séquence complémentaire.

20 La région d'environ 1,1Kb localisée en amont du site d'initiation de la transcription du gène ABCA7, et comprenant le promoteur de base et de multiples éléments régulateurs de la transcription est comprise également dans la séquence identifiée comme la SEQ ID N°2 selon l'invention.

Plus précisément, le nucléotide en position 1 de la séquence SEQ ID N°2 est 25 le nucléotide en position -1111, par rapport au site d'initiation de la transcription du gène ABCA7.

Selon un second aspect, l'invention est relative à un acide nucléique comprenant un polynucléotide ayant au moins 20 nucléotides consécutifs de la 30 séquence nucléotidique SEQ ID N°2 , ou un acide nucléique de séquence complémentaire.

Comme déjà précisé ci-dessus, l'acide nucléique régulateur de la transcription du gène ABCA7 de séquence SEQ ID N°1 comprend, outre une région 5' régulatrice non transcrrite, également la partie 5' du premier exon du gène ABCA7 humain.

La séquence partielle du premier exon du gène ABCA7 est définie comme la séquence SEQ ID N°3.

Selon un troisième aspect, l'invention est relative à un acide nucléique comprenant un polynucléotide ayant au moins 20 nucléotides consécutifs de séquence nucléotidique SEQ ID N° 3 ou un acide nucléique de séquence complémentaire.

De préférence, un acide nucléique selon l'invention sera sous une forme isolée et/ou purifiée.

Fait également partie de l'invention tout fragment "biologiquement actif" d'un acide nucléique tel que défini ci-dessus.

10 Selon encore un autre aspect, l'invention concerne un acide nucléique ayant au moins 80% d'identité en nucléotides avec un acide nucléique tel que défini ci-dessus.

15 En particulier, cet acide nucléique peut être d'origine murine, et est constitué d'un polynucléotide de séquence nucléotidique SEQ ID NO : 4 comprenant de l'extrémité 5' vers 3' :

- une région non transcrrite d'environ 1,2 Kb localisé en amont du site d'initiation de la transcription du gène ABCA7 murin, et
- la séquence partielle du premier exon du gène ABCA7.

20 La région d'environ 1,2Kb localisée en amont du site d'initiation de la transcription du gène ABCA7, et comprenant le promoteur de base et de multiples éléments régulateurs de la transcription est comprise également dans la séquence identifiée comme la SEQ ID NO :5 selon l'invention.

25 L'invention englobe également un acide nucléique caractérisé en ce qu'il hybride, dans des conditions de forte stringence, avec l'un quelconque des acides nucléiques selon l'invention.

30 L'invention concerne aussi un acide nucléique ayant au moins 80%, avantageusement 90%, de préférence 95% et de manière tout à fait préférée 98% d'identité en nucléotides avec un acide nucléique comprenant au moins 20 nucléotides consécutifs d'un polynucléotide choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°5.

ANALYSE DETAILLEE DES SEQUENCES SEQ ID N° :2 ET SEQ ID N° :5

Selon une caractéristique principale, l'acide nucléique de séquence SEQ ID N°2, compris dans l'acide nucléique régulateur du gène humain ABCA7 de séquence SEQ ID N°1, comprend les éléments constitutifs d'un promoteur de base, respectivement une boîte "TATA" dégénérée (TTAAG) localisée 30pb en amont du point d'initiation de la transcription. De même, une boîte "TATA" dégénérée (TTAAA) a été localisée 30pb en amont du point d'initiation de la transcription, sur l'acide nucléique murin de séquence SEQ ID NO : 5, compris dans l'acide nucléique régulateur du gène murin ABCA7 de séquence SEQ ID NO :4. Les boîtes "TATA" sur les promoteurs des gènes humain et murin ABCA7 ainsi que la position des points d'initiation de la transcription sont représentés sur la Figure 1.

Les séquences régulatrices SEQ ID N°2 et SEQ ID N°5 comprennent également de nombreux sites de fixation de divers facteurs de transcription susceptibles de réguler positivement ou négativement l'activité du promoteur de base.

Ainsi, les différentes séquences caractéristiques des sites de fixation de divers facteurs de transcription dans les séquences SEQ ID N°2 et SEQ ID N°5 ont été identifiées par les inventeurs de la manière détaillée ci-après.

Les séquences SEQ ID N°2 et SEQ ID N°5 ont été utilisées comme séquences de référence et traitées selon les algorithmes des logiciels MatInspector (Quandt et al., *Nucl Acid Res* (1995) 23(23), 4878-4884) et comparées aux données répertoriées dans plusieurs bases de données telles que Transfac et la présence ainsi que la localisation des différentes sites caractéristiques des séquences SEQ ID N°2 et 5, et particulièrement les sites de liaison des facteurs de transcription ont été déterminés selon des méthodes bien connues de l'homme du métier.

Plus particulièrement, une analyse détaillée a été réalisée en utilisant les logiciels NNPP (Reese et al. *J. Comput Biol.* (1997) 4(3) 311-23), TSSG, et TSSW (Solorzene et al., *Ismb* (1995), 5, 294-302), sur les 1,1 kb et 1,2Kb en amont du site d'initiation respectivement des séquences SEQ ID N°2 et 5, a permis d'identifier 193 et 233 sites putatifs de liaison aux facteurs de transcription, chez l'homme et la souris lors de la première étape de la recherche. Ceux-ci sont répertoriés dans les tableaux 1 et 2.

Après compilation et filtrage comme décrit ci-dessus, et comparaison des séquences régulatrices humaine et murine, deux modules communs aux séquences régulatrices humaine et murine ont été déterminés, et 5 et 3 sites de fixation possibles de facteurs de transcription différents ont été retenus respectivement dans les modules 1 et 2 sur

les séquences humaine et murine. La position avec les scores de filtration des sites de fixation aux facteurs de transcription identifiés dans les 1111 pb de la séquence SEQ ID N°2 selon l'invention, ainsi que dans les 1220 pb de la séquence SEQ ID NO : 5 selon l'invention, sont présentés dans le tableau 3. Les différents sites de fixation ont 5 été aussi schématiquement représentés dans la figure 1.

Les positions des nucléotides de début de chacun des sites de fixation aux facteurs de transcription sont désignées en référence à la numérotation des nucléotides des séquences SEQ ID N°2 et NO :5 par rapport au point +1 d'initiation de la transcription, contenu dans les séquences SEQ ID N°1 et N°4, telle que représentée 10 sur la figure 1.

La Figure 2 représente la séquence SEQ ID NO : 1, qui contient la séquence SEQ ID N°2. Le premier nucléotide en position 5' de la séquence de la Figure 2 est également le premier nucléotide en position 5' de l'une des séquences nucléiques SEQ ID N°1 et SEQ ID N° 2. Sur la Figure 2, les sites de fixation aux facteurs de 15 transcription sont illustrés en caractères gras qui délimitent leurs positions respectives, et leurs désignations respectives sont indiquées au-dessus de chacune des boîtes correspondantes. La numérotation des nucléotides de la séquence représentée à la figure 2 a été effectuée par rapport au site d'initiation de la transcription, numéroté " +1 ", le nucléotide en 5' du nucléotide +1 étant lui-même numéroté " -1 ".

La Figure 3 représente la séquence SEQ ID NO : 4, qui contient la séquence SEQ ID N°5. Le premier nucléotide en position 5' de la séquence de la figure 3 est également le premier nucléotide en position 5' de l'une des séquences nucléiques SEQ ID N°4 et SEQ ID N° 5. Sur la figure 3, les sites de fixation aux facteurs de transcription sont illustrés en caractères gras qui délimitent leurs positions respectives, et leurs 25 désignations respectives sont indiquées au-dessus de chacune des boîtes correspondantes. La numérotation des nucléotides de la séquence représentée à la figure 3 a été effectuée par rapport au site d'initiation de la transcription, numéroté " +1 ", le nucléotide en 5' du nucléotide +1 étant lui-même numéroté " -1 ".

L'analyse génomique des acides nucléiques régulateurs des séquences 30 humaine et murine SEQ ID NO :2 et 5, a révélé deux modules de régulation qui sont notés module 1 et module 2, et sont particulièrement conservés chez l'homme et la souris. Ces deux modules de régulation comprennent des sites de fixation de facteur de transcription ubiquitaires, tels que NF1, NFY et AP4, ainsi que des sites de fixation

de facteurs de transcription spécifiques du foie tels que CEBP et HNF3B. Ceci est compatible avec les données expérimentales d'expression, présentées dans l'Exemple 3 ci-après, et fournies par Kaminski et al. (Supra), qui montrent une expression du gène ABCA7 dans les tissus fœtaux hépatiques humains.

5 Les deux modules de régulation conservés chez la souris et l'homme comprennent également des sites de fixation de facteurs de transcription tels que GFI1 et NFkappaB (NFkB), qui sont essentiellement présents dans les organes lymphatiques.

10 La description des caractéristiques des sites de fixation à chacun des facteurs de transcription désignés dans les figures 2 et 3 ainsi que dans le tableau 3 peut être aisément retrouvée par l'homme du métier. Une courte description de certains d'entre eux en est faite ci-après.

Facteur NF1:

15 Les caractéristiques de fixation du facteur NF1 peuvent être retrouvées notamment dans les entrées suivantes de la base de données Medline: 88319941, 91219459, 86140112, 87237877, 90174951, 89282387, 90151633, 892618136, 86274639, 87064414, 89263791. Le facteur NF1 reconnaît la séquence palindromique suivante: " TGGCANNNTGCCA (NNTTGGCNNNNNNNNCCNN) " qui est présente 20 dans des promoteurs viraux et cellulaires et au niveau de l'origine de réplication des adénovirus de type 2. Ces protéines sont capables d'activer la transcription et la réplication. Elles se fixent à l'ADN sous la forme d'un homodimère.

Facteur NFY :

25 Le facteur NFY est notamment décrit dans l'entrée n°P25.208 de la base de données Swissprot. C'est un facteur qui reconnaît un motif " CCAAT " dans les séquences promotrices telles que celles du gène codant pour le collagène de type 1, de l'albumine et de la β -actine. Il s'agit d'un stimulateur de la transcription.

Facteur AP4:

L'homme du métier pourra se référer avantageusement aux articles correspondant aux entrées suivantes de la base de données Medline : 2123466,

2833704, 8530024. Le facteur AP4 a un domaine de liaison à l'ADN du type "helix loop helix" (bHLH) ainsi que deux domaines de dimérisation. Le site consensus du facteur AP4 est le suivant « CWCAGCTGGN », et celui-ci se chevauche généralement avec un site de fixation du facteur AP1.

5

CEBP

Les caractéristiques de liaison au facteur CEBP peuvent être retrouvées notamment dans les entrées suivantes de la base de données Medline: 93315489, 91248826, 94193722, 93211931, 92390404, 90258863, 94088523, 90269225 et 10 96133958. Il s'agit d'un activateur de la transcription important dans la régulation de gènes impliqués dans les réponses immunes et inflammatoires. Il se fixe spécifiquement à un élément de réponse à IL-1 dans le gène de l'IL-6. Il joue vraisemblablement un rôle dans la régulation de la phase aiguë de l'inflammation et dans l'hémopoïèse. Le site de reconnaissance consensus est le suivant: "T(T/G) 15 NNGNAA(T/G)".

Facteur HNF3B:

L'homme du métier pourra se référer avantageusement à l'article de OVERDIER et al. (1994, Mol. Cell Biol. vol.14: 2755-2766), ainsi qu'aux entrées suivantes de la base de données Medline: 91352065, 91032994, 92345837, 89160814, 91187609, 91160974, 20 91029477, 94301798 et 94218249. Ce facteur de transcription agit comme activateur de nombreux gènes du foie tels que des gènes AFT, de l'albumine, de la tyrosine aminotransférase et interagit avec des régions régulatrices agissant en *cis* de ces gènes.

25

GFI1

Les caractéristiques de liaison au facteur GFI1 peuvent être retrouvées notamment dans les entrées suivantes de la base de données Medline : 10762661, 9931446, 9571157, 9285685, 9070650, et 7789186. Le gène GFI1 code pour une 30 protéine en doigt de zinc impliquée dans la régulation transcriptionnelle et plus particulièrement dans la voie de signalisation de l'interleukine 2. Le site de reconnaissance consensus est le suivant : « NNNNNNAATCANNNGNNNNNN »

Facteur NFkappa-B:

L'homme du métier pourra avantageusement se référer aux articles correspondant aux entrées suivantes de la base de données Medline: 95369245, 91204058, 94280766, 89345587, 93024383, 88248039, 94173892, 91088538, 5 91239561, 91218850, 92390404, 90156535, 93377072, 92097536, 93309429, 93267517, 92037544, 914266911, 91105848 et 95073993. Le facteur NFkappa-B est un hétérodimère constitué d'une première sous-unité de 50 kDa et d'une seconde sous-unité de 65 kDa. Deux hétérodimères peuvent former un tétramère labile. Sa fixation à l'ADN dépend de la présence de zinc (Zn^{++}). Il peut être induit par de 10 nombreux agents tels que le TNF, la PKA ou encore la PKC. C'est un régulateur clef de gènes impliqués dans des réponses à l'infection, à l'inflammation, ainsi qu'au stress.

Une caractéristique essentielle de l'acide nucléique régulateur selon l'invention, et plus particulièrement de la séquence localisée en amont du site d'initiation de la transcription comprise à la fois dans la séquence SEQ ID N°2 et dans 15 la séquence SEQ ID N°5 est la présence de motifs caractéristiques de sites putatifs de liaison à des facteurs de transcription impliqués dans l'expression génique des lymphocytes T, tels que les facteurs de transcription CEBP, NFkB, et GFI1.

GFI1 est un proto-oncogène qui code pour une protéine nucléaire en doigt de zinc impliquée dans la voie de signalisation par les cytokines et dans l'amplification 20 clonale des cellules T (Zweidler-McKay, et al., *Mol. Cell. Biol.* (1996), **16**(8), 4024-4034). Le facteur de transcription GFI1 qui agit comme un répresseur transcriptionnel des gènes qui inhibent l'activation des cellules T et l'oncogénèse. Il est spécifiquement présent dans le thymus, la rate et les lymphocytes T.

Les facteurs de transcription CEBP et NFkappaB qui sont exprimés dans le 25 thymus, la rate et les lymphocytes T, sont bien connus de l'homme du métier et agissent en coopération dans la médiation de l'induction de l'expression des gènes des lymphocytes T (Runch et al., 1994) et des cellules HepB3 (Shimizu et al., *Gene*, (1994) **149**, 305-310).

Les positions des nucléotides de début, par rapport au site d'initiation de la 30 transcription qui sont à -498 et -469 pour les sites CEBP, et à -260 pour le site NFkB, sur le module régulateur humain, et à -787 et -760, pour les sites CEBP, et à -301 pour le site NFkB, montrent que les deux sites de régulation sont plus éloignés dans le promoteur de souris. Toutefois, il est vraisemblable que les deux sites se rapprochent

dans une structure tridimensionnelle afin de permettre la coactivation par les deux facteurs CEBP et NFkB.

La présence de ces sites potentiels de fixation au CEBP et au NFkB de manière conservée chez l'homme et la souris sur les acides nucléiques régulateurs

5 selon l'invention est compatible avec l'observation selon laquelle l'expression du gène codant pour la protéine ABCA7 humaine est prédominante dans les tissus hématopoïétiques et les lymphocytes T, et interviendrait très probablement dans la médiation cellulaire de l'immunité, notamment dans la pathogenèse de l'athérosclérose (Kaminski et al. Supra).

10 Comme déjà mentionné précédemment, l'invention concerne un acide nucléique comprenant un polynucléotide ayant au moins 20 nucléotides consécutifs de l'une des séquences nucléotidiques SEQ ID N°1 ou 2, et SEQ ID N°4 ou 5, ainsi qu'un acide nucléique de séquence complémentaire.

15 Sont englobés dans la définition ci-dessus les acides nucléiques comprenant un ou plusieurs fragments "biologiquement actifs" de l'une des séquences SEQ ID N°1 ou 2, et SEQ ID N°4 ou 5. L'homme du métier peut aisément obtenir des fragments biologiquement actifs de ces séquences, en se référant notamment au tableau 3 ci-dessus ainsi qu'aux figures 2 et 3 dans lesquels les différents motifs caractéristiques de la séquence régulatrice du gène ABCA7 sont présentés. L'homme du métier pourra 20 ainsi obtenir de tels fragments biologiquement actifs par synthèse chimique totale ou partielle des polynucléotides correspondants ou encore en faisant agir des endonucléases de restriction afin d'obtenir des fragments d'ADN recherchés, les sites de restriction présents sur les séquences SEQ ID N°1 à SEQ ID N°5 pouvant être aisément retrouvés à partir de l'information de séquence, à l'aide de logiciels courants 25 de cartographie de restriction tel que GCG version 9.1 module map.

L'obtention de fragments d'acides nucléiques déterminés à l'aide d'endonucléases de restriction est par exemple décrite dans l'ouvrage de SAMBROOK et al., (*Molecular cloning: a laboratory manual*, 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York (1989).

30 L'invention est donc également relative à un acide nucléique tel que défini ci-dessus, capable de moduler la transcription d'un polynucléotide placé sous son contrôle.

Selon un premier mode de réalisation préféré, un fragment biologiquement actif d'un acide nucléique régulateur de la transcription selon l'invention comprend un premier module conservé (module 2) qui comprend le promoteur de base (boîte TATA) allant du nucléotide en position -1 au nucléotide en position -390, par rapport au site 5 d'initiation de la transcription, le premier nucléotide transcrit étant le nucléotide en position 1112 de la séquence nucléotidique SEQ ID N°1, ou le nucléotide en position 1221 de la séquence nucléotidique SE ID NO :4.

Selon un second mode de réalisation, un fragment biologiquement actif d'un acide nucléique régulateur de la transcription selon l'invention comprend les modules 10 1 et 2 conservés (figure 1) du nucléotide en position -1 au nucléotide en position -860, par rapport au site d'initiation de la transcription, le premier nucléotide transcrit étant le nucléotide en position 1112 de la séquence nucléotidique SEQ ID N°1, ou le nucléotide en position 1221 de la séquence nucléotidique SE ID NO :4.

Selon un troisième mode de réalisation, un tel fragment biologiquement actif 15 d'un acide régulateur de la transcription selon l'invention comprend, outre le promoteur de base (core promoter) et les éléments régulateurs proximaux, également d'autres éléments régulateurs tels que les différents sites GFI1, HNF3B, CEBPB, NF1 et s'étend du nucléotide en position -1 au nucléotide en position -1111, par rapport au site 20 d'initiation de la transcription, le premier nucléotide transcrit étant le nucléotide en position 1112 de la séquence nucléotidique SEQ ID N°1, et au nucléotide en position -1220, par rapport au site d'initiation de la transcription, le premier nucléotide transcrit étant le nucléotide en position 1221 de la séquence nucléotidique SEQ ID N°4.

ANALYSE DE L'EXON 1

25 Le demandeur a également identifié les séquences nucléotidiques localisées en aval du site d'initiation de la transcription et correspondant à l'extrémité 5' de l'exon 1, des gènes humain et murin codant pour la protéine ABCA7.

Plus précisément, l'extrémité 5' de l'exon 1, d'une taille de 1210 nucléotides, 30 débute au nucléotide en position 1112 de la séquence SEQ ID N°1 et se termine au nucléotide en position 2322 de la séquence SEQ ID N°1. L'extrémité 5' de l'exon 1 est identifiée comme la séquence SEQ ID N°3 et la séquence complète de l'exon 1 est identifiée comme la séquence SEQ ID N°6.

L'exon 1 contient le début de la phase ouverte de lecture du gène ABCA7 humain, le nucléotide A du codon ATG étant localisé en position 1208 des séquences SEQ ID N° 3 et 6. L'exon 1 code pour le polypeptide de séquence SEQ ID N° 7.

L'exon 1 est susceptible de contenir des éléments de régulation de 5 l'expression du gène ABCA7, notamment des éléments de type enhancer amplificateur et/ou des éléments de type silencer ou répresseur.

En conséquence, un acide nucléique régulateur de la transcription selon l'invention peut également contenir, outre des fragments biologiquement actifs de la séquence SEQ ID N°1, également des fragments nucléotidiques, voire la totalité des 10 séquences SEQ ID N°2 à SEQ ID N°3 et 6.

Les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°3, et 6, ainsi que leurs fragments, peuvent être notamment utilisés comme sondes ou amorces nucléotidiques pour détecter la présence d'au moins une copie du gène ABCA7 dans un échantillon, ou encore pour amplifier une séquence cible déterminée au sein de la séquence 15 régulatrice du gène ABCA7.

L'invention a donc encore pour objet un acide nucléique ayant au moins 80% d'identité en nucléotides avec un acide nucléique tel que défini ci-dessus, en particulier provenant de l'une des séquences SEQ ID N°1 à SEQ ID N°3 et 6.

L'invention concerne également un acide nucléique hybride, dans des 20 conditions de forte stringence, avec l'un quelconque des acides nucléiques selon l'invention, en particulier un acide nucléique provenant d'une séquence choisie parmi les séquences SEQ ID N°1 à SEQ ID N°3 et 6.

L'invention a également trait à un acide nucléique tel que défini ci-dessus et caractérisé en outre en ce qu'il est capable de moduler la transcription d'un 25 polynucléotide d'intérêt placé sous son contrôle.

Selon un premier aspect, un tel acide nucléique est capable d'activer la transcription du polynucléotide d'intérêt placé sous son contrôle.

Selon un second aspect, un acide nucléique régulateur selon l'invention peut être caractérisé en ce qu'il est capable d'inhiber la transcription du polynucléotide 30 d'intérêt placé sous le contrôle de celui-ci.

Préférentiellement, un acide nucléique régulateur de la transcription selon l'invention, lorsqu'il est localisé convenablement par rapport à un polynucléotide

d'intérêt dont l'expression est recherchée, va permettre la transcription dudit polynucléotide d'intérêt, soit de manière constitutive, soit de manière inductible.

Le caractère inductible de la transcription initiée par un acide nucléique régulateur selon l'invention peut être conféré par un ou plusieurs des éléments 5 régulateurs qu'il contient, par exemple la présence d'un ou plusieurs sites tels que précédemment définis dans la séquence SEQ ID N°1 ou SEQ ID N° 2.

De plus, une expression spécifique de tissu du polynucléotide d'intérêt peut être recherchée en plaçant ce polynucléotide d'intérêt sous le contrôle d'un acide nucléique régulateur selon l'invention capable, par exemple, d'initier la transcription de 10 ce polynucléotide d'intérêt spécifiquement dans certaines catégories de cellules, par exemple des cellules du tissu hématopoïétique, telles les leucocytes périphériques, les cellules du thymus, les cellules de la rate, et la moelle osseuse.

De manière préférée, un acide nucléique régulateur selon l'invention peut comprendre un ou plusieurs éléments régulateurs "discrets", tels que des éléments 15 enhancers et silencers. En particulier, un tel acide nucléique régulateur peut comprendre un ou plusieurs sites de fixation potentiels aux facteurs de transcription tels que définis dans la figure 2.

Un acide régulateur selon l'invention englobe également une séquence ne comprenant pas le promoteur de base, c'est-à-dire la séquence allant du nucléotide en 20 position -1 au nucléotide en position -25, par rapport au site d'initiation de la transcription.

Un tel acide nucléique régulateur comprendra alors préférentiellement un promoteur de base dit "hétérologue", c'est-à-dire un polynucléotide comprenant une boîte "TATA" et une "homéo-boîte" ne provenant pas de l'acide nucléique régulateur 25 du gène ABCA7.

Fait également partie de l'invention, un acide nucléique régulateur de la transcription comprenant tout ou partie de la séquence SEQ ID N°1 qui a été modifié, par exemple par addition, délétion ou substitution d'un ou plusieurs nucléotides. De telles modifications peuvent moduler l'activité transcriptionnelle en provoquant une 30 augmentation ou au contraire une diminution de l'activité du promoteur ou de l'élément régulateur.

Une telle modification peut également affecter la spécificité tissulaire du promoteur ou de l'élément régulateur. Ainsi, par exemple, un acide nucléique

régulateur selon l'invention peut être modifié afin de stimuler la transcription dans seulement l'un des tissus dans lequel il est naturellement exprimé.

Un acide régulateur de la transcription selon l'invention peut également être modifié et être rendu inducible par un composé particulier, par exemple en créant dans 5 la séquence un site inducible par un composé thérapeutique donné.

Les modifications dans une séquence comprenant tout ou partie de la séquence SEQ ID N°1 et comprenant le promoteur ou un élément régulateur peuvent être réalisées avec des méthodes bien connues de l'homme du métier, telle que la mutagenèse. L'activité du promoteur ou de l'élément régulateur modifié peut ensuite 10 être testée, par exemple par clonage du promoteur modifié en amont d'un gène rapporteur, en transfectant la construction d'ADN résultante dans une cellule hôte et en mesurant le niveau d'expression du gène rapporteur dans la cellule hôte transfectée. L'activité du promoteur modifié peut également être analysée *in vivo* dans des animaux transgéniques. Il est également possible de construire des banques de fragments 15 modifiés qui peuvent être ciblées en utilisant des tests fonctionnels dans lesquels, par exemple, seuls les promoteurs ou les éléments régulateurs ayant l'activité désirée seront sélectionnés.

De tels essais peuvent être basés, par exemple sur l'utilisation de gènes rapporteurs conférant une résistance à des composés déterminés, par exemple à des 20 antibiotiques. La sélection de cellules ayant une construction acide nucléique régulateur/gène rapporteur et contenant un promoteur ou un élément régulateur ayant la modification recherchée peut alors être isolée par culture des cellules hôtes transformées avec une telle construction en présence du composé déterminé, par exemple de l'antibiotique déterminé.

25 Le gène rapporteur peut également coder pour toute protéine facilement détectable, par exemple une protéine optiquement détectable telle que la luciférase.

En conséquence, l'invention a également pour objet un acide nucléique comprenant:

- a) un acide nucléique régulateur de la transcription telle que définie ci-dessus;
- 30 et
- b) un polynucléotide d'intérêt codant pour un polypeptide ou un acide nucléique d'intérêt.

Selon un premier aspect, le polynucléotide d'intérêt dont la transcription est recherchée code pour une protéine ou un peptide. La protéine peut être de toute nature, par exemple une protéine d'intérêt thérapeutique, y compris des cytokines, des protéines de structure, des récepteurs ou encore des facteurs de transcription. Par 5 exemple, dans le cas où une transcription spécifiquement dans certains tissus est recherchée, comme par exemple dans des cellules du tissu hématopoïétique, c'est à dire de la rate, de la moelle osseuse, ou dans les leucocytes périphériques, l'acide nucléique régulateur de la transcription comprendra avantageusement un acide nucléique allant du nucléotide en position -1 au nucléotide en position -1111, par 10 rapport au site d'initiation de la transcription de la séquence SEQ ID N°1 ou 2, et allant du nucléotide en position -1 au nucléotide en position -1220 SEQ ID N°4 ou 5.

Dans ce cas, le polynucléotide d'intérêt codera pour un gène impliqué dans la lutte contre l'inflammation, tel qu'un récepteur aux cytokines ou encore pour une superoxyde dismutase. Si un effet antitumoral est recherché, on cherchera alors à 15 stimuler le nombre et l'activation de lymphocytes T cytotoxiques spécifiques d'un antigène tumoral donné.

Dans un autre mode de réalisation, un acide nucléique régulateur selon l'invention sera utilisé en combinaison avec un polynucléotide d'intérêt codant pour la protéine ABCA7.

20 Également, le polynucléotide d'intérêt peut également être un oligonucléotide de type sens.

Comme déjà mentionné, le polynucléotide d'intérêt peut également produire un acide nucléique, tel qu'un acide nucléique antisens spécifique d'un gène dont on cherche à inhiber la traduction.

25

Selon un autre aspect, le polynucléotide d'intérêt dont la transcription est régulée par l'acide nucléique régulateur est un gène rapporteur, tel que tout gène codant pour une protéine détectable.

30 Parmi les gènes rapporteurs préférés, on peut citer notamment le gène de la luciférase, de la β -galactosidase (LacZ), de la chloramphénicol acétyl transférase (CAT) ou encore tout gène codant pour une protéine conférant une résistance à un composé particulier, particulièrement à un antibiotique.

VECTEURS RECOMBINANTS

Par "vecteur" au sens de la présente invention on entendra une molécule d'ADN ou d'ARN circulaire ou linéaire qui est indifféremment sous forme de simple brin ou double brin.

5 Selon un premier mode de réalisation, un vecteur recombinant selon l'invention est utilisé afin d'amplifier l'acide nucléique régulateur selon l'invention qui y est inséré après transformation ou transfection de l'hôte cellulaire désiré.

10 Selon un second mode de réalisation, il s'agit de vecteurs d'expression comprenant, outre un acide nucléique régulateur conforme à l'invention, des séquences dont l'expression est recherchée dans une cellule hôte ou dans un organisme multicellulaire déterminé.

15 Selon un mode de réalisation avantageux, un vecteur recombinant selon l'invention comprendra notamment les éléments suivants :

- (1) un acide nucléique régulateur selon l'invention;
- 15 (2) un polynucléotide d'intérêt comprenant une séquence codante comprise dans l'acide nucléique à insérer dans un tel vecteur, ladite séquence codante étant placée en phase avec les signaux de régulation décrits au (1) ; et
- (3) des séquences d'initiation et d'arrêt de la transcription appropriées.

20 En outre, les vecteurs recombinants selon l'invention pourront inclure une ou plusieurs origines de réplication chez les hôtes cellulaires dans lesquels leur amplification ou leur expression est recherchée, des marqueurs ou des marqueurs de sélection.

25 A titre d'exemples, les promoteurs bactériens pourront être les promoteurs LacI, LacZ, les promoteurs de l'ARN polymérase du bactériophage T3 ou T7, les promoteurs PR, ou PL du phage lambda.

Les promoteurs pour cellules eucaryotes comprendront le promoteur de la thymidine kinase du virus HSV ou encore le promoteur de la métallothionéine-L de souris.

30 De manière générale, pour le choix d'un promoteur adapté, l'homme du métier pourra avantageusement se référer à l'ouvrage de Sambrook et al. (1989) précité ou encore aux techniques décrites par Fuller et al. (1996; *Immunology in Current Protocols in Molecular Biology*).

Lorsque l'expression de la séquence génomique du gène ABCA7 sera désirée, on aura préférentiellement recours à des vecteurs capables d'inclure de grandes séquences d'insertion. Dans ce mode de réalisation particulier, on utilisera de préférence des vecteurs de bactériophages, tels que les vecteurs de bactériophage P1 5 comme le vecteur p158 ou encore le vecteur p158/neo8 décrits par Sternberg (Trends Genet., (1992) 8 : 1-16; *Mamm. Genome* (1994) 5 : 397-404).

Les vecteurs bactériens préférés selon l'invention sont par exemple les vecteurs pBR322(ATCC37017) ou encore des vecteurs tels que pAA223-3 (Pharmacia, Uppsala, Suède), et pGEM1 (Promega Biotech, Madison, WI, ETATS-UNIS).

10 On peut encore citer d'autres vecteurs commercialisés tels que les vecteurs pQE70, pQE60, pQE9 (Qiagen), psiX174, pBluescript SA, pNH8A, pNH16A, pNH18A, pNH46A, pWLNEO, pSV2CAT, pOG44, pXTI, pSG(Stratagene).

Il peut s'agir aussi du vecteur recombinant PXP1 décrit par Nordeen SK et al. (BioTechniques, (1988) 6 : 454-457).

15 Il peut s'agir également de vecteurs de type *Baculovirus* tels que le vecteur pVL1392/1393 (Pharmingen) utilisé pour transfacter les cellules de la lignée Sf9 (ATCC N°CRL 1711) dérivées de *Spodoptera frugiperda*.

Il peut encore s'agir de vecteurs adénoviraux tels que l'adénovirus humain de type 2 ou 5.

20 Un vecteur recombinant selon l'invention peut aussi être un vecteur rétroviral ou encore un vecteur adéno-associé (AAV). De tels vecteurs adéno-associés sont par exemple décrits par Flotte et al., *Am. J. Respir., Cell Mol. Biol.* (1992) 7 : 349-356 ; Samulski et al., *J. Virol.* (1989) 63 : 3822-3828 ; MacLaughlin et al., *Am. J. Hum. Genet.* (1996) 59 : 561-569.

25 Pour permettre l'expression d'un polynucléotide d'intérêt sous le contrôle d'un acide nucléique régulateur selon l'invention, la construction polynucléotidique comprenant la séquence régulatrice et la séquence codante doit être introduite dans une cellule hôte. L'introduction d'une telle construction polynucléotidique selon l'invention dans une cellule hôte peut être réalisée *in vitro*, selon les techniques bien 30 connues de l'homme du métier pour transformer ou transfacter des cellules, soit en culture primaire, soit sous la forme de lignées cellulaires. On peut aussi réaliser

l'introduction des polynucléotides selon l'invention *in vivo* ou *ex vivo*, pour la prévention ou le traitement de maladies liées à un déficit dans le transport en protéine ABCA7.

Pour introduire les polynucléotides ou les vecteurs dans une cellule hôte, l'homme du métier pourra avantageusement se référer à différentes techniques, 5 comme la technique de précipitation au phosphate de calcium (Graham et al., *Virology* (1973) **52** : 456-457 ; Chen et al., *Mol. Cell. Biol.* (1987) **7** : 2745-2752), le DEAE Dextran (Gopal et al., *Mol. Cell. Biol.*, (1985) **5** : 1188-1190), l'électroporation (Tur-Kaspa et al., *Mol. Cell. Biol.*, (1986) **6** : 716-718.; Potter et al., *Proc. Natl. Acad. Sci. USA* (1984), **81**(22), 7161-5), la microinjection directe (Harland et al., *J Cell Biol* (1985) **101** : 10 1094-1095), les liposomes chargés en ADN (Nicolau et al., *Methods Enzymol* (1987) **149** : 157-76 ; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) **76** : 3348-3352).

Une fois que le polynucléotide a été introduit dans la cellule hôte, il peut être intégré de manière stable dans le génome de la cellule. L'intégration peut être réalisée à un endroit précis du génome, par recombinaison homologue, ou peut être intégré au 15 hasard. Dans certains modes de réalisation, le polynucléotide peut être maintenu de manière stable dans la cellule hôte sous la forme d'un fragment d'épisome, l'épisome comprenant des séquences permettant le maintien et la réPLICATION de ce dernier, soit de manière indépendante, soit de manière synchronisée avec le cycle cellulaire.

Selon un mode de réalisation particulier, une méthode pour introduire un 20 polynucléotide selon l'invention dans une cellule hôte, en particulier une cellule hôte provenant d'un mammifère, *in vivo*, comprend une étape au cours de laquelle on introduit une préparation comprenant un vecteur d'expression pharmaceutiquement compatible et un polynucléotide "nu" selon l'invention, placé sous le contrôle de séquences de régulation appropriées, par injection locale au niveau du tissus choisi, 25 par exemple un tissu musculaire lisse, le polynucléotide "nu" étant absorbé par les cellules de ce tissu.

Des compositions pour l'utilisation *in vitro* et *in vivo* comprenant des polynucléotides "nus" sont par exemples décrites dans la demande PCT N° WO 95/11307 ainsi que dans les articles de Tacson et al. (*Nature Med.* (1996) **2**(8), 888-30 892) et de Huygen et al., (*Nat. Med.* (1996) **2**(8), 893-898).

Selon un mode de réalisation spécifique de l'invention, il est fourni une composition pour la production *in vivo* d'une protéine d'intérêt. Cette composition

comprend un polynucléotide codant pour le polypeptide d'intérêt placé sous le contrôle d'une séquence régulatrice selon l'invention, en solution dans un vecteur physiologiquement acceptable.

La quantité de vecteur qui est injectée à l'organisme hôte choisi varie selon le 5 site de l'injection. A titre indicatif, il peut être injecté entre environ 0,1 et environ 100 µg de la construction polynucléotidique séquence régulatrice/séquence codante dans le corps d'un animal.

Lorsque l'acide nucléique régulateur selon l'invention est localisé, sur la construction polynucléotidique (ou vecteur), de telle manière à contrôler la transcription 10 d'une séquence comprenant un cadre de lecture ouvert codant la protéine ABCA7, le vecteur est injecté de préférence dans le corps d'un patient susceptible de développer une maladie liée à un déficit en protéine ABCA7.

En conséquence, l'invention concerne également une composition pharmaceutique destinée à la prévention ou au traitement de sujets affectés d'un 15 dysfonctionnement de la protéine ABCA7, comprenant un acide nucléique régulateur selon l'invention et un polynucléotide d'intérêt codant pour la protéine ABCA7, en association avec un ou plusieurs excipients physiologiquement compatibles.

Avantageusement, une telle composition comprendra l'acide nucléique régulateur défini par l'une des séquences SEQ ID N°1 ou 2, et SEQ ID N°4 ou 5, ou un 20 fragment biologiquement actif de cet acide nucléique régulateur.

L'invention a en outre pour objet une composition pharmaceutique destinée à la prévention ou au traitement de sujets affectés d'un dysfonctionnement dans le métabolisme des lipides, comprenant un vecteur recombinant tel que défini ci-dessus, en association avec un ou plusieurs excipients physiologiquement compatibles.

25 L'invention a également pour objet une composition pharmaceutique destinée à la prévention ou au traitement de sujets affectés d'un dysfonctionnement des processus impliquant le système immunitaire et l'inflammation, comprenant un vecteur recombinant tel que défini ci-dessus, en association avec un ou plusieurs excipients physiologiquement compatibles.

30 L'invention est également relative à l'utilisation d'une construction polynucléotidique conforme à l'invention et comprenant un acide nucléique régulateur du gène ABCA7 ainsi qu'une séquence codant pour la protéine ABCA7, pour la fabrication d'un médicament destiné à la prévention ou au traitement de sujets affectés

d'un dysfonctionnement du métabolisme des lipides ou d'un problème d'origine immunologique ou encore d'origine inflammatoire.

L'invention a également trait à l'utilisation d'un vecteur recombinant selon l'invention, comprenant, outre un acide nucléique régulateur de l'invention, un acide nucléique codant pour la protéine ABCA7, pour la fabrication d'un médicament destiné à la prévention ou au traitement de sujets affectés d'un dysfonctionnement des processus impliquant le système immunitaire et l'inflammation.

10 *Vecteurs utiles dans des procédés de thérapie génique somatique et compositions contenant de tels vecteurs*

La présente invention concerne aussi une nouvelle approche thérapeutique pour le traitement et/ou la prévention des pathologies liées au métabolisme des lipides ainsi que pour le traitement et/ou la prévention des pathologies liées au dysfonctionnement des mécanismes de médiation lymphocytaire de l'inflammation. Elle propose une solution avantageuse aux inconvénients de l'art antérieur, en démontrant la possibilité de traiter les pathologies, notamment des pathologies liées à un dysfonctionnement du métabolisme des lipides dans les tissus myélo-lymphatiques, par la thérapie génique, par le transfert et l'expression *in vivo* d'une construction polynucléotidique comprenant, outre un acide nucléique régulateur selon l'invention, une séquence codant pour une protéine ABCA7 qui est fortement présumée comme étant impliquée dans le transport et/ou le métabolisme des lipides. L'invention offre ainsi un moyen simple permettant un traitement spécifique et efficace des sujets affectés d'un dysfonctionnement des processus impliquant le système immunitaire et l'inflammation.

25 La thérapie génique consiste à corriger une déficience ou une anomalie (mutation, expression aberrante, etc.) ou à assurer l'expression d'une protéine d'intérêt thérapeutique par introduction d'une information génétique dans la cellule ou l'organe affecté. Cette information génétique peut être introduite soit *ex vivo* dans une cellule extraite de l'organe, la cellule modifiée étant alors réintroduite dans l'organisme, soit 30 directement *in vivo* dans le tissu approprié. Dans ce second cas, différentes techniques existent, parmi lesquelles des techniques diverses de transfection impliquant des complexes d'ADN et de DEAE-dextran (Pagano et al., J.Virol. 1 (1967) 891), d'ADN et

de protéines nucléaires (Kaneda et al., *Science* 243 (1989) 375), d'ADN et de lipides (Felgner et al., *PNAS* 84 (1987) 7413), l'emploi de liposomes (Fraley et al., *J.Biol.Chem.* 255 (1980) 10431), etc. Plus récemment, l'emploi de virus comme vecteurs pour le transfert de gènes est apparu comme une alternative prometteuse à 5 ces techniques physiques de transfection. A cet égard, différents virus ont été testés pour leur capacité à infecter certaines populations cellulaires. En particulier, les rétrovirus (RSV, HMS, MMS, etc.), le virus HSV, les virus adéno-associés, et les adénovirus.

La présente invention est donc également relative à une nouvelle approche 10 thérapeutique pour le traitement des pathologies liées au transport des lipides, consistant à transférer et à exprimer *in vivo* des gènes codant pour ABCA7 placés sous le contrôle d'un acide régulateur selon l'invention. Il est particulièrement avantageux de construire des virus recombinants contenant une séquence d'ADN comprenant un acide nucléique régulateur selon l'invention et une séquence codant pour une protéine 15 ABCA7 impliquée dans le métabolisme de lipides, d'administrer ces virus recombinants *in vivo*, et que *cette* administration permette une expression stable et efficace d'une protéine ABCA7 biologiquement active *in vivo*, et sans effet cytopathologique.

Les adénovirus constituent des vecteurs particulièrement efficaces pour le transfert et l'expression du gène ABCA7. En particulier, l'utilisation d'adénovirus 20 recombinants comme vecteurs permet d'obtenir des niveaux d'expression suffisamment élevés du gène d'intérêt pour produire l'effet thérapeutique recherché. D'autres vecteurs viraux, tels que les rétrovirus ou les virus adéno-associés (AAV), permettant une expression stable du gène sont aussi revendiqués.

La présente invention offre ainsi une nouvelle approche pour le traitement et la 25 prévention des pathologies liées à des dysfonctionnements dans le métabolisme des lipides et dans les voies de signalisation de l'inflammation par les lymphocytes.

L'invention a donc aussi pour objet un virus recombinant défectif comprenant un acide nucléique régulateur selon l'invention et une séquence nucléique codant pour une protéine ABCA7 impliquée dans le métabolisme des lipides ou dans les processus 30 impliquant le système immunitaire et l'inflammation.

L'invention a également trait à l'utilisation d'un tel virus recombinant défectif pour la préparation d'une composition pharmaceutique destinée au traitement et/ou à la

prévention des dysfonctionnements de la signalisation de l'inflammation par les lymphocytes.

La présente invention concerne également l'utilisation de cellules modifiées génétiquement *ex vivo* par un virus tel que décrit ci-dessus, ou de cellules productrices de tels virus, implantées dans l'organisme, permettant une expression *in vivo* prolongée et efficace d'une protéine ABCA7 biologiquement active.

La présente invention montre qu'il est possible d'incorporer une séquence d'ADN codant pour ABCA7 dans le contrôle d'un acide nucléique régulateur tel que défini ci-dessus dans un vecteur viral, et que ces vecteurs permettent d'exprimer efficacement une forme mature, biologiquement active. Plus particulièrement, l'invention montre que l'expression *in vivo* de ABCA7 peut être obtenue par administration directe d'un adénovirus ou par implantation d'une cellule productrice ou génétiquement modifiée par un adénovirus ou par un rétrovirus incorporant un tel ADN.

La présente invention est particulièrement avantageuse, car elle permet d'induire une expression contrôlée et sans effet nocif de ABCA7 dans des organes qui ne sont pas normalement concernés par l'expression de cette protéine. En particulier, une libération significative de la protéine ABCA7 est obtenue par implantation de cellules productrices de vecteurs de l'invention, ou infectées *ex vivo* par des vecteurs de l'invention.

L'activité de médiateur dans le métabolisme des lipides produit dans le cadre de la présente invention peut être du type ABCA7 humaine ou animale. La séquence nucléique utilisée dans le cadre de la présente invention peut être un ADNc, un ADN génomique (ADNg), un ARN (dans le cas des rétrovirus) ou une construction hybride consistant par exemple en un ADNc dans lequel seraient insérés un ou plusieurs introns. Il peut également s'agir de séquences synthétiques ou semi-synthétiques. De manière particulièrement avantageuse, on utilise un ADNc ou un ADNg. En particulier, l'utilisation d'un ADNg permet une meilleure expression dans les cellules humaines. Pour permettre leur incorporation dans un vecteur viral selon l'invention, ces séquences sont avantageusement modifiées, par exemple par mutagenèse dirigée, en particulier pour l'insertion de sites de restriction appropriés. Les séquences décrites dans l'art antérieur ne sont en effet pas construites pour une utilisation selon l'invention, et des adaptations préalables peuvent s'avérer nécessaires, pour obtenir des expressions importantes. Dans le cadre de la présente invention, on préfère utiliser une séquence

nucléique codant pour une protéine ABCA7 humaine. Par ailleurs, il est également possible d'utiliser une construction codant pour un dérivé de ces protéines ABCA7. Un dérivé de ces protéines ABCA7 comprend par exemple toute séquence obtenue par mutation, délétion et/ou addition par rapport à la séquence native, et codant pour un

5 produit conservant l'activité de médiateur du métabolisme des lipides. Ces modifications peuvent être réalisées par les techniques connues de l'homme du métier (voir techniques générales de biologie moléculaire ci-après). L'activité biologique des dérivés ainsi obtenus peut ensuite être aisément déterminée, comme indiqué notamment dans les exemples décrivant la mesure de l'efflux de lipides à partir des

10 cellules. Les dérivés au sens de l'invention peuvent également être obtenus par hybridation à partir de banques d'acides nucléiques, en utilisant comme sonde la séquence native ou un fragment de celle-ci.

Ces dérivés sont notamment des molécules ayant une plus grande affinité pour leurs sites de fixation, des molécules présentant une plus grande résistance aux

15 protéases, des molécules ayant une efficacité thérapeutique plus grande ou des effets secondaires moindres, ou éventuellement ayant de nouvelles propriétés biologiques. Les dérivés incluent également les séquences d'ADN modifiées permettant une expression améliorée *in vivo*.

Dans un premier mode de réalisation, la présente invention concerne un virus

20 recombinant défectif comprenant un acide nucléique régulateur selon l'invention et une séquence d'ADNc codant pour une protéine ABCA7 impliquée dans le transport et le métabolisme du cholestérol. Dans un autre mode de réalisation préféré de l'invention, la séquence d'ADN est une séquence d'ADNg. La séquence d'ADNc codant pour la

25 protéine ABCA7, et utilisable dans un vecteur selon l'invention est avantageusement la séquence SEQ ID N°8.

Les vecteurs de l'invention peuvent être préparés à partir de différents types de virus. Préférentiellement, on utilise des vecteurs dérivés des adénovirus, des virus adéno-associés (AAV), des virus de l'herpès (HSV) ou des rétrovirus. Il est tout particulièrement avantageux d'utiliser un adénovirus, pour une administration directe ou

30 pour la modification *ex vivo* de cellules destinées à être implantées, ou un rétrovirus, pour l'implantation de cellules productrices.

Les virus selon l'invention sont défectifs, c'est-à-dire qu'ils sont incapables de se répliquer de façon autonome dans la cellule cible. Généralement, le génome des

virus défectifs utilisés dans le cadre de la présente invention est donc dépourvu au moins des séquences nécessaires à la réplication dudit virus dans la cellule infectée. Ces régions peuvent être soit éliminées (en tout ou en partie), soit rendues non fonctionnelles, soit substituées par d'autres séquences et notamment par la séquence 5 nucléique codant pour la protéine ABCA7. Préférentiellement, le virus défectif conserve néanmoins les séquences de son génome qui sont nécessaires à l'encapsidation des particules virales.

S'agissant plus particulièrement d'adénovirus, différents sérotypes, dont la structure et les propriétés varient quelque peu, ont été caractérisés. Parmi ces 10 sérotypes, on préfère utiliser dans le cadre de la présente invention les adénovirus humains de type 2 ou 5 (Ad 2 ou Ad 5) ou les adénovirus d'origine animale (voir demande WO 94/26914). Parmi les adénovirus d'origine animale utilisables dans le cadre de la présente invention on peut citer les adénovirus d'origine canine, bovine, murine, (exemple : Mav1, Beard et al., *Virology* 75 (1990) 81), ovine, porcine, aviaire ou 15 encore simienne (exemple : SAV). De préférence, l'adénovirus d'origine animale est un adénovirus canin, plus préférentiellement un adénovirus CAV2 [souche Manhattan ou A26/61 (ATCC VR-800) par exemple]. De préférence, on utilise dans le cadre de l'invention des adénovirus d'origine humaine ou canine ou mixte. Préférentiellement, les adénovirus défectifs de l'invention comprennent les ITR, une séquence permettant 20 l'encapsidation et la séquence codant pour la protéine ABCA7 placée sous le contrôle d'un acide nucléique selon l'invention. Avantageusement, dans le génome des adénovirus de l'invention, la région E1 au moins est rendue non fonctionnelle. Encore plus préférentiellement, dans le génome des adénovirus de l'invention, le gène E1 et au moins l'un des gènes E2, E4, L1-L5 sont non fonctionnels. Le gène viral considéré peut 25 être rendu non fonctionnel par toute technique connue de l'homme du métier, et notamment par suppression totale, substitution, délétion partielle, ou addition d'une ou plusieurs bases dans le ou les gènes considérés. De telles modifications peuvent être obtenues *in vitro* (sur de l'ADN isolé) ou *in situ*, par exemple, au moyens des techniques du génie génétique, ou encore par traitement au moyen d'agents 30 mutagènes. D'autres régions peuvent également être modifiées, et notamment la région E3 (WO 95/02697), E2 (WO 94/28938), E4 (WO 94/28152, WO 94/12649, WO 95/02697) et L5 (WO 95/02697). Selon un mode préféré de mise en œuvre, l'adénovirus selon l'invention comprend une délétion dans les régions E1 et E4 et la

séquence codant pour ABCA7 est insérée au niveau de la région E1 inactivée. Selon un autre mode de réalisation préféré, il comprend une délétion dans la région E1 au niveau de laquelle sont insérés la région E4 et la séquence codant pour ABCA7 (Demande de brevet Français FR 94 13355).

5 Les adénovirus recombinants défectifs selon l'invention peuvent être préparés par toute technique connue de l'homme du métier (Levrero et al., *Gene* (1991) **101**: 195, EP 185 573; Graham, *EMBO J.* (1984) **3**: 2917). En particulier, ils peuvent être préparés par recombinaison homologue entre un adénovirus et un plasmide portant entre autre la séquence d'ADN codant pour la protéine ABCA7. La recombinaison 10 homologue se produit après co-transfection desdits adénovirus et plasmide dans une lignée cellulaire appropriée. La lignée cellulaire utilisée doit de préférence (i) être transformable par lesdits éléments, et (ii), comporter les séquences capables de complémer la partie du génome de l'adénovirus défectif, de préférence sous forme intégrée pour éviter les risques de recombinaison. A titre d'exemple de lignée, on peut 15 mentionner la lignée de rein embryonnaire humain 293 (Graham et al., *J. Gen. Virol.* (1977) **36** :59) qui contient notamment, intégrée dans son génome, la partie gauche du génome d'un adénovirus Ad5 (12 %) ou des lignées capables de complémer les fonctions E1 et E4 telles que décrites notamment dans les demandes n° WO 94/26914 et WO 95/02697.

20 Ensuite, les adénovirus qui se sont multipliés sont récupérés et purifiés selon les techniques classiques de biologie moléculaire, comme illustré dans les exemples.

Concernant les virus adéno-associés (AAV), il s'agit de virus à ADN de taille relativement réduite, qui s'intègrent dans le génome des cellules qu'ils infectent, de manière stable et site-spécifique. Ils sont capables d'infecter un large spectre de 25 cellules, sans induire d'effet sur la croissance, la morphologie ou la différenciation cellulaire. Par ailleurs, ils ne semblent pas impliqués dans des pathologies chez l'homme. Le génome des AAV a été cloné, séquencé et caractérisé. Il comprend environ 4700 bases, et contient à chaque extrémité une région répétée inversée (ITR) de 145 bases environ, servant d'origine de réplication pour le virus. Le reste du génome 30 est divisé en 2 régions essentielles portant les fonctions d'encapsidation : la partie gauche du génome, qui contient le gène rep impliqué dans la réplication virale et l'expression des gènes viraux; la partie droite du génome, qui contient le gène cap codant pour les protéines de capsid du virus.

L'utilisation de vecteurs dérivés des AAV pour le transfert de gènes in vitro et in vivo a été décrite dans la littérature (voir notamment WO 91/18088; WO 93/09239; US 4,797,368, US 5,139,941, EP 488 528). Ces documents décrivent différentes constructions dérivées des AAV, dans lesquelles les gènes rep et/ou cap sont déletés et remplacés par un gène d'intérêt, et leur utilisation pour transférer in vitro (sur cellules en culture) ou in vivo (directement dans un organisme) ledit gène d'intérêt. Toutefois, aucun de ces documents ne décrit ni ne suggère l'utilisation d'un AAV recombinant pour le transfert et l'expression in vivo ou ex vivo d'une protéine ABCA7, ni les avantages d'un tel transfert. Les AAV recombinants défectifs selon l'invention peuvent 5 être préparés par co-transfection, dans une lignée cellulaire infectée par un virus auxiliaire humain (par exemple un adénovirus), d'un plasmide contenant la séquence codant pour la protéine ABCA7 bordée de deux régions répétées inversées (ITR) d'AAV, et d'un plasmide portant les gènes d'encapsidation (gènes rep et cap) d'AAV. Les AAV recombinants produits sont ensuite purifiés par des techniques classiques.

10

15 Concernant les virus de l'herpès et les rétrovirus, la construction de vecteurs recombinants a été largement décrite dans la littérature : voir notamment Breakfield et al., (New Biologist 3 (1991) 203); EP 453242, EP 178220, Bernstein et al. (Genet. Eng. 7 (1985) 235); McCormick, (BioTechnology 3 (1985) 689), etc.

20 En particulier, les rétrovirus sont des virus intégratifs, infectant les cellules en division. Le génome des rétrovirus comprend essentiellement deux LTR, une séquence d'encapsidation et trois régions codantes (gag, pol et env). Dans les vecteurs recombinants dérivés des rétrovirus, les gènes gag, pol et env sont généralement déletés, en tout ou en partie, et remplacés par une séquence d'acide nucléique hétérologue d'intérêt. Ces vecteurs peuvent être réalisés à partir de 25 différents types de rétrovirus tels que notamment le MoMuLV ("murine moloney leukemia virus"; encore désigné MoMLV), le MSV ("murine moloney sarcoma virus"), le HaSV ("harvey sarcoma virus"); le SNV ("spleen necrosis virus"); le RSV ("rous sarcoma virus") ou encore le virus de Friend.

30 Pour construire des rétrovirus recombinants comportant une séquence codant pour la protéine ABCA7 placée sous le contrôle d'un acide nucléique régulateur selon l'invention, un plasmide comportant notamment les LTR, la séquence d'encapsidation et ladite séquence codante est généralement construit, puis utilisé pour transfecter une lignée cellulaire dite d'encapsidation, capable d'apporter en trans les fonctions

rétrrovirales déficientes dans le plasmide. Généralement, les lignées d'encapsidation sont donc capables d'exprimer les gènes gag, pol et env. De telles lignées d'encapsidation ont été décrites dans l'art antérieur, et notamment la lignée PA317 (US 4,861,719); la lignée PsiCRIP (WO 90/02806) et la lignée GP+envAm-12 (WO 89/07150). Par ailleurs, les rétrrovirus recombinants peuvent comporter des modifications au niveau des LTR pour supprimer l'activité transcriptionnelle, ainsi que des séquences d'encapsidation étendues, comportant une partie du gène gag (Bender et al., J. Virol. 61 (1987) 1639). Les rétrrovirus recombinants produits sont ensuite purifiés par des techniques classiques.

10 Pour la mise en œuvre de la présente invention, il est tout particulièrement avantageux d'utiliser un adénovirus recombinant défectif. Les résultats donnés ci-après démontrent en effet les propriétés particulièrement intéressantes des adénovirus pour l'expression *in vivo* d'une protéine ayant une activité de médiateur du métabolisme des lipides. Les vecteurs adénoviraux selon l'invention sont particulièrement avantageux 15 pour une administration directe *in vivo* d'une suspension purifiée, ou pour la transformation *ex vivo* de cellules, notamment autologues, en vue de leur implantation. De plus, les vecteurs adénoviraux selon l'invention présentent en outre des avantages importants, tels que notamment leur très haute efficacité d'infection, permettant de réaliser des infections à partir de faibles volumes de suspension virale.

20 Selon un autre mode particulièrement avantageux de mise en œuvre de l'invention, on utilise une lignée productrice de vecteurs rétrroviraux contenant un acide nucléique régulateur selon l'invention et la séquence codant pour la protéine ABCA7, pour une implantation *in vivo*. Les lignées utilisables à cet effet sont notamment les cellules PA317 (US4,861,719), PsiCrip (WO 90/02806) et GP+envAm-12 (US 25 5,278,056), modifiées pour permettre la production d'un rétrrovirus contenant une séquence nucléique codant pour une protéine ABCA7 selon l'invention. Par exemple des cellules souches totipotentes, précurseurs des lignées cellulaires sanguines, peuvent être prélevées et isolées chez le sujet. Ces cellules mises en culture peuvent être alors transfectées par le vecteur rétrroviral contenant la séquence codant pour la 30 protéine ABCA7 sous le contrôle de son propre promoteur. Ces cellules sont alors réintroduites chez le sujet. La différentiation de ces cellules sera à l'origine de cellules du tissu hématopoïétique exprimant la protéine ABCA7, notamment des lymphocytes T qui, participent à la signalisation de l'inflammation.

Avantageusement, dans les vecteurs de l'invention, la séquence codant pour la protéine ABCA7 est placée sous le contrôle d'un acide régulateur selon l'invention comprenant les éléments régulateurs permettant son expression dans les cellules infectées, et tout particulièrement les éléments régulateurs de type NFkappaB, CEBP, 5 et GFI1.

Comme indiqué ci-dessus, la présente invention concerne également toute utilisation d'un virus tel que décrit ci-dessus pour la préparation d'une composition pharmaceutique destinée au traitement et/ou à la prévention des pathologies liées au métabolisme des lipides ou encore au dysfonctionnement lié aux processus impliquant 10 le système immunitaire et l'inflammation.

La présente invention concerne également une composition pharmaceutique comprenant un ou plusieurs virus recombinants défectifs tels que décrits précédemment. Ces compositions pharmaceutiques peuvent être formulées en vue d'administrations par voie topique, orale, parentérale, intranasale, intraveineuse, 15 intramusculaire, sous-cutanée, intra-oculaire, transdermique, etc. De préférence, les compositions pharmaceutiques de l'invention contiennent un véhicule pharmaceutiquement acceptable pour une formulation injectable, notamment pour une injection intraveineuse, telle que par exemple dans la veine porte du patient. Il peut s'agir en particulier de solutions stériles, isotoniques, ou de compositions sèches, 20 notamment lyophilisées, qui, par addition selon le cas d'eau stérilisée ou de sérum physiologique, permettent la constitution de solutés injectables. L'injection directe dans la veine porte du patient est avantageuse car elle permet de cibler l'infection au niveau du foie et ainsi, de concentrer l'effet thérapeutique au niveau de cet organe.

Les doses de virus recombinant défectif utilisées pour l'injection peuvent être 25 adaptées en fonction de différents paramètres, et notamment en fonction du vecteur viral, du mode d'administration utilisé, de la pathologie concernée ou encore de la durée du traitement recherchée. D'une manière générale, les adénovirus recombinants selon l'invention sont formulés et administrés sous forme de doses comprises entre 10^4 et 10^{14} pfu/ml, et de préférence 10^6 à 10^{10} pfu/ml. Le terme pfu ("plaque forming 30 unit") correspond au pouvoir infectieux d'une solution de virus, et est déterminé par infection d'une culture cellulaire appropriée, et mesure, généralement après 48 heures, du nombre de plages de cellules infectées. Les techniques de détermination du titre pfu d'une solution virale sont bien documentées dans la littérature.

Concernant les rétrovirus, les compositions selon l'invention peuvent comporter directement les cellules productrices, en vue de leur implantation.

A cet égard, un autre objet de l'invention concerne toute cellule de mammifère infectée par un ou plusieurs virus recombinants défectifs tels que décrits ci-dessus.

5 Plus particulièrement, l'invention concerne toute population de cellules humaines infectées par ces virus. Il peut s'agir en particulier de cellules d'origine sanguine (cellules souches totipotentes ou précurseurs), fibroblastes, myoblastes, hépatocytes, kératinocytes, cellules musculaires lisses et endothéliales, cellules gliales, etc.

Les cellules selon l'invention peuvent être issues de cultures primaires. Celles-
10 ci peuvent être prélevées par toute technique connue de l'homme du métier, puis mises en culture dans des conditions permettant leur prolifération. S'agissant plus particulièrement de fibroblastes, ceux-ci peuvent être aisément obtenus à partir de biopsies, par exemple selon la technique décrite par Ham (*Methods Cell Biol* (1980) 21a: 255). Ces cellules peuvent être utilisées directement pour l'infection par les virus,
15 ou conservées, par exemple par congélation, pour l'établissement de banques autologues, en vue d'une utilisation ultérieure. Les cellules selon l'invention peuvent également être des cultures secondaires, obtenues par exemple à partir de banques préétablies (voir par exemple EP 228458, EP 289034, EP 400047, EP 456640).

Les cellules en culture sont ensuite infectées par les virus recombinants, pour
20 leur conférer la capacité de produire une protéine ABCA7 biologiquement active. L'infection est réalisée *in vitro* selon des techniques connues de l'homme du métier. En particulier, selon le type de cellules utilisé et le nombre de copies de virus par cellule désiré, l'homme du métier peut adapter la multiplicité d'infections et éventuellement le nombre de cycles d'infections réalisé. Il est bien entendu que ces étapes doivent être
25 effectuées dans des conditions de stérilité appropriées lorsque les cellules sont destinées à une administration *in vivo*. Les doses de virus recombinant utilisées pour l'infection des cellules peuvent être adaptées par l'homme du métier selon le but recherché. Les conditions décrites ci-avant pour l'administration *in vivo* peuvent être appliquées à l'infection *in vitro*. Pour l'infection par des rétrovirus, il est également
30 possible de co-cultiver les cellules que l'on désire infecter avec des cellules productrices des rétrovirus recombinants selon l'invention. Ceci permet de s'affranchir de la purification des rétrovirus.

Un autre objet de l'invention concerne un implant comprenant des cellules de mammifères infectées par un ou plusieurs virus recombinants défectifs telles que décrites ci-dessus ou des cellules productrices de virus recombinants, et une matrice extracellulaire. Préférentiellement, les implants selon l'invention comprennent 10⁵ à 5 10¹⁰ cellules. Plus préférentiellement, ils en comprennent 10⁶ à 10⁸.

Plus particulièrement, dans les implants de l'invention, la matrice extracellulaire comprend un composé gélifiant et éventuellement un support permettant l'ancrage des cellules.

10 Pour la préparation des implants selon l'invention, différents types de gélifiants peuvent être employés. Les gélifiants sont utilisés pour l'inclusion des cellules dans une matrice ayant la constitution d'un gel, et pour favoriser l'ancrage des cellules sur le support, le cas échéant. Différents agents d'adhésion cellulaire peuvent donc être utilisés comme gélifiants, tels que notamment le collagène, la gélatine, les glycosaminoglycans, la fibronectine, les lectines, etc. De préférence, on utilise dans le 15 cadre de la présente invention du collagène. Il peut s'agir de collagène d'origine humaine, bovine ou murine. Plus préférentiellement, on utilise du collagène de type I.

20 Comme indiqué ci avant, les compositions selon l'invention comprennent avantageusement un support permettant l'ancrage des cellules. Le terme ancrage désigne toute forme d'interaction biologique et/ou chimique et/ou physique entraînant l'adhésion et/ou la fixation des cellules sur le support. Par ailleurs, les cellules peuvent soit recouvrir le support utilisé, soit pénétrer à l'intérieur de ce support, soit les deux. On préfère utiliser dans le cadre de l'invention un support solide, non toxique et/ou biocompatible. En particulier, on peut utiliser des fibres de polytétrafluoroéthylène (PTFE) ou un support d'origine biologique.

25 La présente invention offre ainsi un moyen très efficace pour le traitement ou la prévention des pathologies liées au transport du cholestérol, en particulier l'obésité, l'hypertriglycéridémie, ou, dans le domaine des affections cardio-vasculaires, l'infarctus du myocarde, l'angor, la mort subite, la décompensation cardiaque et les accidents cérébro-vasculaires.

30 En outre, ce traitement peut concerner aussi bien l'homme que tout animal tel que les ovins, les bovins, les animaux domestiques (chiens, chats, etc.), les chevaux, les poissons, etc.

CELLULES HOTES RECOMBINANTES

L'invention concerne aussi une cellule hôte recombinante comprenant l'un quelconque des acides nucléiques de l'invention de séquence SEQ ID N° 1 à SEQ ID N°6, et plus particulièrement un acide nucléique de séquence SEQ ID NO 1 à SEQ ID N°3.

Selon un autre aspect, l'invention est également relative à une cellule hôte recombinante comprenant un vecteur recombinant tel que ci-dessus décrit.

Les cellules hôtes préférées selon l'invention sont par exemple les suivantes :

10 a) cellules hôtes procaryotes: souches d'*Escherichia coli* (souche DH5- α), de *Bacillus subtilis*, de *Salmonella typhimurium*, ou encore des souches d'espèces telles que *Pseudomonas*, *Streptomyces* et *Staphylococcus* ;

15 b) cellules hôtes eucaryotes: cellules HeLa (ATCC N°CCL2), cellules Cv 1 (ATCC N°CCL70), cellules COS (ATCC N°CRL 1650), cellules Sf-9 (ATCC N°CRL 1711), cellules CHO (ATCC N°CCL-61) ou encore cellules 3T3 (ATCC N°CRL-6361), ou encore les cellules de la lignée Hepa1-6 référencées à l'American Type Culture Collection (ATCC, Rockville, MD, Etats-Unis d'Amérique).

20 c) les cellules en culture primaire provenant d'un individu chez lequel l'expression d'un acide nucléique d'intérêt placé sous le contrôle d'un acide nucléique régulateur selon l'invention est recherchée.

d) Les cellules à multiplication indéfinie (lignées cellulaires) obtenues à partir des cellules en culture primaire du c) ci-dessus, selon des techniques bien connues de l'homme du métier.

25

PROCEDES DE CRIBLAGE

Procédé de criblage *in vitro*.

30 L'invention fournit des procédés pour traiter un sujet affecté d'une pathologie liée au niveau d'expression de la protéine ABCA7. En particulier, une telle méthode de traitement consiste à administrer au sujet un composé modulant l'expression du gène

ABCA7, qui peut être identifié selon divers procédés de criblage *in vitro* tels que définis ci-après.

Une première méthode consiste à identifier des composés modulant l'expression du gène ABCA7. Selon une telle méthode, des cellules exprimant le gène ABCA7 sont incubées avec une substance ou molécule candidate à tester et le niveau d'expression de l'ARN messager de ABCA7 ou encore le niveau de production de la protéine ABCA7 est ensuite déterminé.

Les niveaux d'ARN messager de ABCA7 peuvent être déterminés par hybridation sur gel de type Northern, bien connu de l'homme du métier. Les niveaux d'ARN messager de ABCA7 peuvent également être déterminés par des procédés mettant en oeuvre la PCR ou encore la technique décrite par WEBB et al. (*Journal of Biomolecular Screening* (1996), vol. 1: 119).

Les niveaux de production de la protéine ABCA7 peuvent être déterminés par immunoprecipitation ou immunochimie en utilisant un anticorps qui reconnaît spécifiquement la protéine ABCA7.

Selon une autre méthode de criblage d'une molécule ou substance candidate modulant l'activité d'un acide nucléique régulateur selon l'invention, une construction nucléotidique telle que définie ci-dessus, comprenant un acide nucléique régulateur selon l'invention ainsi qu'un polynucléotide rapporteur placé sous le contrôle de l'acide nucléique régulateur, est utilisée, ledit acide nucléique régulateur comprenant au moins un promoteur de base et au moins un élément régulateur de l'une des séquences SEQ ID N°1 à SEQ ID N°3. Le polynucléotide rapporteur peut être un gène codant pour une protéine détectable, tel qu'un gène codant pour une luciférase.

Selon un tel procédé de criblage, les cellules sont transfectées avec la construction polynucléotidique contenant l'acide nucléique régulateur selon l'invention et le polynucléotide rapporteur, de manière stable ou transitoire.

Les cellules transformées sont ensuite incubées en présence ou en l'absence de la molécule ou de la substance candidate à tester pendant un temps suffisant, puis le niveau d'expression du gène rapporteur est déterminé. Les composés qui induisent un changement statistiquement significatif de l'expression du gène rapporteur (soit une augmentation, soit au contraire une diminution de l'expression du gène rapporteur) sont ensuite identifiés et, le cas échéant, sélectionnés.

Ainsi, l'invention a encore pour objet un procédé pour le criblage *in vitro* d'une molécule ou d'une substance modulant l'activité d'un acide nucléique régulateur selon l'invention, notamment modulant la transcription du polynucléotide rapporteur constitutif d'une construction polynucléotidique selon l'invention, caractérisé en ce qu'il comporte 5 les étapes consistant à :

- a) mettre en culture une cellule hôte recombinante comprenant un polynucléotide d'intérêt placé sous le contrôle d'un acide nucléique régulateur selon l'invention;
- 10 b) incuber la cellule hôte recombinante avec la substance ou molécule à tester;
- c) détecter l'expression du polynucléotide d'intérêt;
- d) comparer les résultats obtenus à l'étape c) avec les résultats obtenus lorsque la cellule hôte recombinante est mise en culture en l'absence de la molécule ou substance candidate à tester.

15 L'invention concerne aussi un kit ou nécessaire pour le criblage *in vitro* d'une molécule ou d'une substance candidate capable de moduler l'activité d'un acide nucléique régulateur selon l'invention, comprenant:

- 20 a) une cellule hôte transformée avec une construction polynucléotidique telle que définie ci-dessus, comprenant un polynucléotide rapporteur d'intérêt placé sous le contrôle d'un acide nucléique régulateur selon l'invention; et
- b) le cas échéant, des moyens de détection de l'expression du polynucléotide rapporteur d'intérêt.

25 De préférence, le polynucléotide rapporteur d'intérêt est la séquence codante d'une luciférase. Dans ce cas, l'acide nucléique régulateur selon l'invention est inséré dans un vecteur, en amont de la séquence codant pour la luciférase. Il peut s'agir par exemple du vecteur pGL3-basic (pGL3-b) commercialisé par la Société PROMEGA (Madison, Wisconsin, Etats-Unis).

30 Dans ce cas, le vecteur recombinant comprenant la séquence codante de la luciférase placée sous le contrôle d'un acide nucléique régulateur selon l'invention est transfété dans les cellules hôtes recombinantes dont l'activité luciférase est ensuite déterminée après mise en culture en présence ou non de la substance ou de la molécule candidate à tester.

On peut dans ce cas utiliser comme témoins des vecteurs pGL3-b contenant soit le promoteur du cytomégalovirus (CMV), le promoteur ApoAI ou encore aucun promoteur. Pour le test d'activité luciférase, les cellules transfectées sont lavées avec un tampon PBS et lysées avec 500 µl de tampon de lyse (50 mM tris, 150 mM NaCl, 5 0,02% d'azide de sodium, 1% de NP-40, 100 µg/ml de AEBSF et 5 µg/ml de leupeptine).

5 50 µl du lysat cellulaire obtenu sont ensuite ajoutés à 100 µl du substrat de la luciférase (Promega) et les mesures d'activité sont réalisées sur un lecteur spectrophotométrique de microplaqué, 5 minutes après l'addition du lysat cellulaire.

10 Les données sont exprimées en unités relatives d'activité luciférase. Les constructions polynucléotidiques produisant de hauts niveaux d'activité luciférase dans les cellules transfectées sont celles qui contiennent un acide nucléique régulateur selon l'invention compris dans la séquence SEQ ID N°1 qui est capable de stimuler la transcription.

15 Pour les mesures des niveaux d'expression d'ARN messager dans un procédé de criblage selon l'invention, on prépare tout d'abord des sondes spécifiques de l'ARN messager du polynucléotide rapporteur d'intérêt, par exemple à l'aide du kit multiprime labeling kit (commercialisé par la Société Amersham Life Sciences, Cleveland, Ohio, Etats-Unis).

20

PROCEDE DE CRIBLAGE IN VIVO

Selon un autre aspect de l'invention, des compositions modulant l'activité d'un acide nucléique régulateur selon l'invention peuvent être identifiées *in vivo*, dans des animaux transgéniques non humains.

25 Selon une telle méthode, un animal transgénique non humain, par exemple une souris, est traité avec une molécule ou une substance candidate à tester, par exemple une substance ou molécule candidate auparavant sélectionnée par un procédé de criblage *in vitro* tel que défini ci-dessus.

30 Après une durée déterminée, le niveau d'activité de l'acide nucléique régulateur selon l'invention est déterminé et comparé à l'activité d'un animal transgénique non humain identique, par exemple une souris transgénique identique, qui n'a pas reçu la molécule ou la substance candidate.

L'activité de l'acide nucléique régulateur selon l'invention fonctionnel dans l'animal transgénique peut être déterminée par diverses méthodes, par exemple la mesure des niveaux d'ARN messager correspondant aux polynucléotides rapporteurs d'intérêt placés sous le contrôle dudit acide nucléique régulateur par hybridation de type Northern, ou encore par hybridation *in situ*. ou encore par imagerie biophotonique non invasive (Xenogen Corporation).

5 Selon une alternative, l'activité de l'acide nucléique régulateur selon l'invention peut être déterminée en mesurant les niveaux d'expression de protéine codée par les polynucléotides rapporteurs d'intérêt, par exemple par immunohistochimie, dans le cas 10 où le polynucléotide rapporteur d'intérêt comprend un cadre de lecture ouvert codant pour une protéine détectable par une telle technique.

15 Pour la mise en oeuvre d'un procédé de criblage *in vivo* d'une substance ou molécule candidate modulant l'activité d'un acide nucléique régulateur selon l'invention, on préférera des mammifères non humains tels que des souris, des rats ou des 20 cobayes ou des lapins qui ont leur génome modifié par l'insertion d'une construction polynucléotidique comprenant un polynucléotide rapporteur d'intérêt placé sous le contrôle d'un acide nucléique régulateur selon l'invention.

Les animaux transgéniques selon l'invention comprennent le transgène, c'est-25 à-dire la construction polynucléotidique précitée; dans une pluralité de leurs cellules somatiques et/ou germinales.

La construction d'animaux transgéniques selon l'invention peut être réalisée selon des techniques classiques bien connues de l'homme du métier. L'homme du métier pourra en particulier se référer à la production d'animaux transgéniques, et 30 particulièrement à la production de souris transgéniques, telles que décrites dans les brevets US N°4,873,191 (délivré le 10 Octobre 1989), US N°5,464,764 (délivré le 7 Novembre 1995) et US 5,789,215 (délivré le 4 Août 1998) le contenu de ces documents étant ici incorporé par référence.

En bref, une construction polynucléotidique comprenant un acide nucléique régulateur selon l'invention et un polynucléotide rapporteur d'intérêt placé sous le 35 contrôle de ce dernier est insérée dans une lignée de cellules souches de type ES. L'insertion de la construction polynucléotidique est réalisée de préférence par électroporation, telle que décrite par Thomas et al. (1987, Cell, Vol.51:503-512).

Les cellules ayant subi l'étape d'électroporation sont ensuite criblées pour la présence de la construction polynucléotidique (par exemple par sélection à l'aide de marqueurs, ou encore par PCR ou par analyse sur gel d'électrophorèse d'ADN de type Southern) afin de sélectionner les cellules positives ayant intégré la construction 5 polynucléotidique exogène dans leur génome, le cas échéant à la suite d'un événement de recombinaison homologue. Une telle technique est par exemple décrite par MANSOUR et al. (*Nature* (1988) 336: 348-352).

Ensuite, les cellules sélectionnées positivement sont isolées, clonées et injectées dans des blastocystes de souris de 3,5 jours, comme cela est décrit par 10 BRADLEY (1987, *Production and Analysis of Chimaeric mice*. In: E. J. ROBERTSON (Ed., teratocarcinomas and embryonic stem cells: A practical approach. IRL press, Oxford, page 113). Des blastocystes sont ensuite introduits dans un animal hôte femelle et le développement de l'embryon est poursuivi jusqu'au terme.

Selon une alternative, des cellules de type ES sélectionnées positivement sont 15 mises en contact avec des embryons de 2,5 jours à un stade 8-16 cellules (*morulae*) comme décrit par WOOD et al. (1993, *Proc. Natl. Acad. Sci. USA*, vol.90: 4582-4585) ou par NAGY et al. (1993, *Proc. Natl. Acad. Sci. USA*, vol. 90: 8424-8428), les cellules ES étant internalisées afin de coloniser extensivement le blastocyste, y compris les cellules donnant naissance à la lignée germinale.

20 Les descendants sont ensuite testés afin de déterminer ceux qui ont intégré la construction polynucléotidique (le transgène).

L'invention a donc également pour objet un animal transgénique non humain dont les cellules somatiques et/ou germinales ont été transformées par un acide nucléique ou une construction polynucléotidique selon l'invention.

25 L'invention a également trait à des cellules hôtes recombinantes obtenues à partir d'un animal transgénique tel que décrit ci-dessus. Des lignées de cellules recombinantes provenant d'un animal transgénique selon l'invention peuvent être établies en culture à long terme à partir de n'importe quel tissu d'un tel animal transgénique, par exemple par transfection des cultures de cellules primaires avec des 30 vecteurs exprimant des oncogènes tels que le grand antigène T de SV40, tel que décrit par exemple par CHOU (1989, *Mol. Endocrinol.* Vol.3: 1511-1514) et SCHAY et al. (1991, *Biochem. Biophys. Acta*, vol. 1072: 1-7).

L'invention est également relative à un procédé pour le criblage *in vivo* d'une molécule ou d'une substance candidate modulant l'activité d'un acide nucléique régulateur selon l'invention, comportant les étapes consistant à a) administrer la substance ou la molécule candidate à un animal transgénique tel que défini ci-dessus ;

5 b) détecter le niveau d'expression d'un polynucléotide rapporteur d'intérêt placé sous le contrôle de l'acide nucléique régulateur ; c) comparer les résultats obtenus au b) avec les résultats obtenus chez un animal transgénique n'ayant pas reçu la substance ou la molécule candidate.

L'invention est en outre relative à un procédé de criblage *in vivo* d'une substance ou molécule modulant la transcription d'un polynucléotide d'intérêt constitutif d'un acide nucléique selon l'invention, caractérisé en ce qu'il comprend les étapes consistant à (a) administrer la substance ou molécule candidate à un mammifère transgénique non humain tel que défini ci-dessus; (b) détecter l'expression du polynucléotide d'intérêt chez le mammifère transgénique tel que traité à l'étape a); et

15 (c) comparer les résultats de détection de l'étape (b) aux résultats observés chez un mammifère transgénique non humain tel que défini ci-dessus n'ayant pas reçu l'administration de la substance ou molécule candidate.

L'invention se rapporte aussi à un kit ou nécessaire pour le criblage *in vivo* d'une molécule ou substance candidate modulant l'activité d'un acide nucléique régulateur selon l'invention, comprenant (a) un animal transgénique tel que défini ci-dessus ; (b) le cas échéant, les moyens de détection du niveau d'expression du polynucléotide rapporteur d'intérêt.

COMPOSES ET COMPOSITIONS PHARMACEUTIQUES

L'invention concerne aussi des compositions pharmaceutiques destinées à la prévention ou au traitement d'un déficit dans le métabolisme des lipides, ou d'un dysfonctionnement dans les processus impliquant le système immunitaire et l'inflammation.

En premier lieu, l'invention a également pour objet une substance ou une molécule candidate modulant l'activité d'un acide nucléique régulateur selon l'invention.

30 De manière tout à fait préférée, l'invention concerne également une substance ou une molécule candidate caractérisée en ce qu'elle augmente l'activité d'un acide nucléique régulateur selon l'invention, et tout particulièrement d'un acide nucléique régulateur comprenant la séquence SEQ ID N°1, 2, 4 ou 5.

Préférentiellement, une telle substance ou molécule capable de moduler l'activité d'un acide nucléique régulateur selon l'invention a été sélectionnée selon l'un des procédés de criblage in vitro ou in vivo défini ci-dessus.

Ainsi, un sujet affecté dans le métabolisme des lipides ou dans la signalisation 5 de l'immunité est traité par l'administration à ce sujet d'une quantité efficace d'un composé modulant l'activité d'un acide nucléique régulateur selon l'invention.

Ainsi, un patient ayant une faible activité du promoteur ABCA7 peut être traité avec une molécule ou une substance précitée afin d'augmenter l'activité du promoteur ABCA7.

10 De manière alternative, un patient ayant une activité du promoteur ABCA7 anormalement haute peut être traité avec un composé capable de diminuer ou de bloquer l'activité du promoteur ABCA7.

Ainsi, la présente invention se rapporte également à une substance ou molécule utilisée en tant que principe actif d'un médicament.

15 Un tel composé peut être un composé qui module l'interaction d'au moins un facteur de transcription avec le promoteur ABCA7 ou un élément régulateur d'un acide nucléique régulateur selon l'invention.

Par exemple, le composé peut inhiber l'interaction de l'un des facteurs de transcription listés dans le tableau 1 avec un acide nucléique régulateur selon 20 l'invention.

Le composé peut également être un composé qui module l'activité d'un facteur de transcription se fixant au promoteur de ABCA7 ou encore d'un élément régulateur présent sur ce dernier.

Un composé d'intérêt thérapeutique selon l'invention peut également être un 25 composé qui module l'interaction d'un premier facteur de transcription avec un second facteur de transcription.

Comme détaillé dans l'analyse des différents facteurs de transcription susceptibles de se fixer à la séquence SEQ ID N°1, 2, 4, ou 5 certains facteurs de transcription sont actifs seulement s'ils sont associés avec un autre facteur de 30 transcription.

Un composé d'intérêt thérapeutique selon l'invention est préférentiellement choisi parmi les acides nucléiques, les peptides et les petites molécules.

Par exemple, un tel composé peut être un acide nucléique antisens qui se fixe

spécifiquement sur une région du promoteur ABCA7 ou sur un élément régulateur d'un acide nucléique régulateur d'ABCA7 et inhibant ou diminuant l'activité du promoteur.

5 Ce composé d'intérêt thérapeutique peut également être un acide nucléique antisens qui interagit spécifiquement avec un gène codant pour un facteur de transcription modulant l'activité du promoteur ABCA7, de telle manière que l'interaction de l'acide nucléique antisens avec le gène codant pour le facteur de transcription se fixant sur le promoteur ABCA7 diminue la production de ce facteur de transcription, résultant en une augmentation ou une diminution de l'activité du promoteur ABCA7, selon que le facteur de transcription augmente ou au contraire réduit l'activité du 10 promoteur ABCA7.

La toxicité et l'efficacité thérapeutique des composés thérapeutiques selon l'invention peuvent être déterminées selon les protocoles pharmaceutiques standard dans des cellules en culture ou chez des animaux expérimentaux, par exemple pour déterminer la dose létale LD₅₀ (c'est-à-dire la dose létale pour 50% de la population 15 testée) ainsi que la dose efficace ED₅₀ (c'est-à-dire la dose thérapeutiquement efficace dans 50% de la population testée).

Pour tous les composés d'intérêt thérapeutique selon l'invention, la dose thérapeutiquement efficace peut être estimée initialement à partir de tests réalisés dans des cultures cellulaires *in vitro*.

20 L'invention a également pour objet des compositions pharmaceutiques comprenant une quantité thérapeutiquement efficace d'une substance ou molécule d'intérêt thérapeutique selon l'invention.

25 Les compositions pharmaceutiques selon l'invention sont plus particulièrement destinées au traitement et/ou à la prévention des déficiences dans le métabolisme des lipides, ou dans les mécanismes impliquant le système immunitaire et l'inflammation.

De telles compositions pharmaceutiques peuvent être formulées de manière classique en utilisant un ou plusieurs vecteurs ou excipients physiologiquement acceptables.

30 Ainsi, les composés d'intérêt thérapeutique selon l'invention, ainsi que leurs sels et solvates physiologiquement acceptables, peuvent être formulés pour une administration par injection, inhalation ou encore par administration orale, buccale, parentérale ou rectale. Des techniques de préparation de compositions pharmaceutiques selon l'invention peuvent être aisément retrouvées par l'homme du

métier, par exemple dans l'ouvrage Remmington's Pharmaceutical Sciences, Mead Publishing Co., Easton, PA, Etats-Unis.

Pour une administration systémique, on préférera l'injection, y compris les injections intramusculaires, intraveineuses, intrapéritonéales et sous-cutanées. Dans ce cas, les compositions pharmaceutiques selon l'invention peuvent être formulées sous forme de solutions liquides, de préférence dans des solutions ou tampons physiologiquement compatibles.

**PROCEDE DE DETECTION D'UNE ALTERATION DE LA TRANSCRIPTION DU
10 GENE ABCA7 HUMAIN**

L'invention a en outre pour objet des procédés pour déterminer si un sujet présente un risque pour le développement d'une pathologie liée à un déficit dans le métabolisme des lipides, ou dans les processus impliquant le système immunitaire et
15 l'inflammation.

De telles méthodes comprennent la détection, dans des cellules d'un échantillon biologique provenant du sujet à tester, de la présence ou de l'absence d'une altération génétique caractérisée par une altération dans l'expression d'un gène dont l'expression est régulée par le promoteur de ABCA7.

20 A titre illustratif, de telles altérations génétiques peuvent être détectées afin de déterminer l'existence d'une délétion d'un ou plusieurs nucléotides dans la séquence d'un acide nucléique régulateur d'ABCA7, de l'addition d'un ou plusieurs nucléotides ou encore de la substitution d'un ou plusieurs nucléotidiques dans ladite séquence SEQ ID N°1, 2, 3 ou 6.

25 Selon un mode particulier de réalisation d'un procédé de détection d'une altération de la transcription du gène ABCA7 chez un sujet, l'altération génétique est identifiée selon une méthode comprenant le séquençage de tout ou partie de la séquence SEQ ID N°1, ou alternativement de tout ou partie d'au moins la séquence SEQ ID N°2.

30 Des amorces de séquençage peuvent être construites afin d'hybrider avec une région déterminée de la séquence SEQ ID N°1. De telles amorces de séquençage sont construites préférentiellement de manière à amplifier des fragments d'environ 300

à environ 500 nucléotides de la séquence SEQ ID N°1 ou d'une séquence complémentaire.

Les fragments amplifiés, par exemple par la méthode PCR, sont ensuite séquencés et la séquence obtenue est comparée à la séquence de référence SEQ ID N°1 afin de déterminer si une ou plusieurs délétions, additions ou substitutions de nucléotides sont retrouvées dans la séquence amplifiée à partir de l'ADN contenu dans l'échantillon biologique provenant du sujet testé.

L'invention concerne donc encore un procédé de détection d'une altération de la transcription du gène ABCA7 chez un sujet, comportant les étapes suivantes :

- 10 a) séquençage d'un fragment d'acide nucléique amplifiable à l'aide d'au moins une amorce nucléotidique hybrideant avec la séquence SEQ ID N°1 ou SEQ ID N°2 selon l'invention;
- 15 b) aligner la séquence obtenue en a) avec la séquence SEQ ID N°1 ou la SEQ ID N°2;
- 15 c) déterminer la présence d'une ou plusieurs délétions, additions ou substitutions d'au moins un nucléotide dans la séquence du fragment d'acide nucléique, par rapport à la séquence SEQ ID N°1 ou SEQ ID N° 2 de référence.

L'invention concerne également un kit ou nécessaire pour la détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant une ou plusieurs amorces de séquençage capables de s'hybrider avec une région de la séquence SEQ ID No : 1, et ainsi de permettre le séquençage d'un polynucléotide localisé en amont du site d'initiation de la transcription du gène ABCA7 chez le sujet à tester.

En outre, font également partie de l'invention des sondes oligonucléotidiques hybrideant avec une région de la séquence SEQ ID N°1 ou de la séquence SEQ ID N° 2 dans laquelle une altération dans la séquence a été déterminée lors de la mise en oeuvre du procédé de détection décrit ci-dessus.

De manière alternative, font aussi partie de l'invention des sondes oligonucléotidiques hybrideant spécifiquement avec une région correspondante de la séquence SEQ ID N°1 ou de la séquence SEQ ID N°2, pour laquelle une ou plusieurs délétions, additions ou substitutions d'au moins un nucléotide a été déterminée chez un sujet.

5 De telles sondes oligonucléotidiques constituent des moyens de détection d'altérations dans la séquence régulatrice du gène ABCA7 et donc également des moyens de détection d'une prédisposition au développement d'une pathologie liée à un déficit dans le métabolisme des lipides ou à dysfonctionnement au niveau des processus impliquant le système immunitaire et l'inflammation.

L'invention a donc encore pour objet un kit ou nécessaire de détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant:

- 10 a) une ou plusieurs amores hybrideant avec une région de la séquence SEQ ID N°1 ou de la séquence SEQ ID N°2 ;
- b) le cas échéant, les moyens nécessaires à la réalisation d'une réaction d'amplification.

L'invention a encore pour objet un kit ou nécessaire pour la détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant :

- 15 a) une ou plusieurs sondes oligonucléotidiques telles que définies ci-dessus;
- b) le cas échéant, les réactifs nécessaires à la réalisation d'une réaction d'hybridation.

20 L'invention a également pour objet un kit ou nécessaire pour la détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant une ou plusieurs sondes hybrideant avec une région de la séquence SEQ ID No : 1 ou de la séquence SEQ ID No : 2 permettant de quantifier l'ARN messager de ABCA7 dans un matériel biologique provenant dudit sujet à tester.

25 Les fragments d'acides nucléiques dérivés de l'une quelconque des séquences nucléotidiques SEQ ID N° 1-6 sont donc utiles pour la détection de la présence d'au moins une copie d'une séquence nucléotidique régulatrice du gène ABCA7 ou encore d'un fragment ou d'un variant (contenant une mutation ou un polymorphisme) de cette dernière dans un échantillon.

Les sondes ou les amores nucléotidiques selon l'invention comprennent au moins 8 nucléotides consécutifs d'un acide nucléique choisi dans le groupe constitué des séquences SEQ ID NO 1-5, ou d'un acide nucléique de séquence complémentaire.

30 De préférence, des sondes ou amores nucléotidiques selon l'invention auront une longueur de 10, 12, 15, 18 ou 20 à 25, 35, 40, 50, 70, 80, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide nucléique selon l'invention, en particulier un

acide nucléique de séquence nucléotidique choisie parmi les séquences SEQ ID NO 1-5.

Alternativement, une sonde ou une amorce nucléotidique selon l'invention consistera et/ou comprendra les fragments d'une longueur de 12, 15, 18, 20, 25, 35, 5 40, 50, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide nucléique selon l'invention, plus particulièrement d'un acide nucléique choisi parmi les séquences SEQ ID N°1-5, ou d'un acide nucléique de séquence complémentaire.

La définition d'une sonde et d'une amorce nucléotidique selon l'invention englobe donc des oligonucléotides qui hybrident, dans les conditions d'hybridation de 10 forte stringence définies ci-avant, avec un acide nucléique choisi parmi les séquences SEQ ID NO 1-5, 6 ou 8 ou avec une séquence complémentaire de ces derniers.

Des exemples d'amorces et de couples d'amorces permettant d'amplifier différentes régions du gène ABCA7 sont représentés ci-dessous.

Il s'agit par exemple du couple d'amorces représenté par l'amorce de 15 séquence SEQ ID N°9 : AGCCAGCAACGCAATCCTCC et l'amorce de séquence SEQ ID N°10: CGCACCATGTCAATGAGCCC.

Une amorce ou une sonde nucléotidique selon l'invention peut être préparée par toute méthode adaptée bien connue de l'homme du métier, y compris par clonage et action d'enzymes de restriction ou encore par synthèse chimique directe selon des 20 techniques telles que la méthode au phosphodiester de Narang et al., (*Methods Enzymol* (1979) **68**:90-98) ou de Brown et al. (*Methods Enzymol* (1979) **68**:109-151), la méthode aux diéthylphosphoramidites de Beaucage et al. (*Tetrahedron Lett* (1980) **22**: 1859-1862) ou encore la technique sur support solide décrite dans le brevet EP 0 707 592.

25 Chacun des acides nucléiques selon l'invention, y compris les sondes et amorces oligonucléotidiques décrites ci-dessus, peut être marqué, si désiré, en incorporant un marqueur détectable par des moyens spectroscopiques, photochimiques, biochimiques, immunochimiques ou encore chimiques.

Par exemple, de tels marqueurs peuvent consister en des isotopes radioactifs 30 (32P, 33P, 3H, 35S), des molécules fluorescentes (5-bromodéoxyuridine, fluorescéine, acétylaminofluorène, digoxigénine) ou encore des ligands tels que la biotine.

Le marquage des sondes est réalisé de préférence par incorporation de molécules marquées au sein des polynucléotides par extension d'amorces, ou bien par rajout sur les extrémités 5' ou 3' ou encore par « nick translation ».

Des exemples de marquage non radioactifs de fragments d'acides nucléiques 5 sont décrits notamment dans le brevet français n° FR 78 109 75 ou encore dans les articles de Urdea et al. (*Nucleic Acid Res (1988) 11: 4937-4957*) ou Sanchez-pescador et al. (*J Clin Microbiol (1988) 26(10) 1934-1938*).

De manière avantageuse, les sondes selon l'invention peuvent avoir des caractéristiques structurelles de nature à permettre une amplification du signal, telles 10 que les sondes décrites par Urdea et al. (*Mol. Cell. Biol., (1991) 6 : 716-718*), ou encore dans le brevet européen n° EP-0 225 807 (CHIRON).

Les sondes oligonucléotides selon l'invention peuvent être utilisées notamment dans des hybridations de type Southern à l'ADN génomique ou bien de type northern à l'ARN.

15 Les sondes selon l'invention peuvent aussi être utilisées pour la détection de produits d'amplification PCR ou encore pour la détection de mésappariements.

Des sondes ou amorces nucléotidiques selon l'invention peuvent être immobilisées sur un support solide. De tels supports solides sont bien connus de l'homme du métier et comprennent des surfaces des puits de plaques de microtitration, 20 des lits de polystyrène, des lits magnétiques, des bandes de nitrocellulose, ou encore des microparticules telles que des particules de latex.

En conséquence, la présente invention concerne également un procédé de détection de la présence d'un acide nucléique tel que décrit ci-dessus dans un échantillon, ladite méthode comprenant les étapes consistant à :

25 1) mettre en contact une ou plusieurs sondes nucléotidiques selon l'invention avec l'échantillon à tester ;

2) détecter le complexe éventuellement formé entre la ou les sondes et l'acide nucléique présent dans l'échantillon.

Selon un mode de réalisation particulier du procédé de détection selon 30 l'invention, la ou les sondes oligonucléotidiques sont immobilisées sur un support.

Selon un autre aspect, les sondes oligonucléotidiques comprennent un marqueur détectable.

L'invention concerne en outre un nécessaire ou kit pour la détection de la présence d'un acide nucléique selon l'invention dans un échantillon, ledit nécessaire comprenant :

5 a) une ou plusieurs sondes nucléotidiques telles que décrites ci-dessus ;
b) le cas échéant, les réactifs nécessaires à la réaction d'hybridation.

Selon un premier aspect, le nécessaire ou kit de détection est caractérisé en ce que la ou les sondes sont immobilisées sur un support.

Selon un second aspect, le nécessaire ou kit de détection est caractérisé en ce que les sondes oligonucléotidiques comprennent un marqueur détectable.

10 Selon un mode de réalisation particulier du kit de détection décrit ci-dessus, un tel kit comprendra une pluralité de sondes oligonucléotidiques conformes à l'invention qui pourront être utilisées pour détecter des séquences cibles d'intérêt ou alternativement détecter des mutations dans les régions codantes ou les régions non codantes des acides nucléiques selon l'invention, plus particulièrement des acides 15 nucléiques de séquences SEQ ID NO 1-5, 6 et 8 ou les acides nucléiques de séquence complémentaire.

Ainsi, les sondes selon l'invention immobilisées sur un support peuvent être ordonnées en matrices telles que les "puces à ADN". De telles matrices ordonnées ont été en particulier décrites dans le brevet US N° 5,143,854, dans les demandes PCT 20 N° WO 90/150 70 et 92/10092.

Des matrices supports sur lesquelles des sondes oligonucléotidiques ont été immobilisées à une haute densité sont par exemple décrites dans les brevets US N°5,412,087 et dans la demande PCT N°WO 95/11995.

25 Les amorces nucléotidiques selon l'invention peuvent être utilisées pour amplifier l'un quelconque des acides nucléiques selon l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences SEQ ID NO 1-5, ou encore un variant de celui-ci.

30 Un autre objet de l'invention concerne un procédé pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement un acide nucléique de séquences SEQ ID NO 1-5 ou un fragment ou un variant de celui-ci contenu dans un échantillon, ledit procédé comprenant les étapes consistant à :

a) mettre en contact l'échantillon, dans lequel la présence de l'acide nucléique cible est suspectée, avec une paire d'amorces nucléotidiques dont la position

d'hybridation est localisée respectivement du côté 5' et du côté 3' de la région de l'acide nucléique cible dont l'amplification est recherchée, en présence des réactifs nécessaires à la réaction d'amplification ; et

b) détecter des acides nucléiques amplifiés.

5 Pour mettre en œuvre le procédé d'amplification tel que défini ci-dessus, on aura avantageusement recours à l'une quelconque des amores nucléotidiques décrites ci-avant.

L'invention a en outre pour objet un nécessaire ou kit pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement tout ou partie d'un acide 10 nucléique de séquences SEQ ID NO 1-5 ledit nécessaire ou kit comprenant :

a) un couple d'amores nucléotidiques conformes à l'invention, dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de l'acide nucléique cible dont l'amplification est recherchée ;

b) le cas échéant, les réactifs nécessaires à la réaction d'amplification.

15 Un tel nécessaire ou kit d'amplification comprendra avantageusement au moins une paire d'amores nucléotidiques telles que décrites ci-dessus.

L'invention est en outre illustrée, sans pour autant être limitée, par les figures et les exemples ci-après.

20 La figure 1 est une représentation schématique des sites de facteurs de transcription retrouvés chez l'homme et la souris au sein de la région promotrice des gènes ABCA7.

La figure 2 illustre la séquence SEQ ID N°1 et la position de chacun des motifs caractéristiques de fixation à différents facteurs de transcription est représentée en caractères gras, la désignation du facteur de transcription spécifique de la séquence 25 correspondante étant indiquée au-dessus de la séquence nucléotidique.

La figure 3 illustre la séquence SEQ ID N°4 et la position de chacun des motifs caractéristiques de fixation à différents facteurs de transcription est représentée en caractères gras, la désignation du facteur de transcription spécifique de la séquence correspondante étant indiquée au-dessus de la séquence nucléotidique.

30 La figure 4 illustre le pattern d'expression du gène ABCA7 humain sur des northern blots de différents tissus adultes et fœtaux (Clontech) hybridés avec un amplimère réalisé avec les amores SEQ ID N°9 et 10 (tableau 4).

La figure 5 illustre le pattern d'expression du gène ABCA7 murin sur un northern blot de différents tissus adultes hybridés avec un amplimère réalisé avec des amorces spécifiques du transcrit murin.

La figure 6 représente le profil d'expression du gène codant pour la protéine ABCA7 sur une coupe transversale d'artère inflammée, par hybridation *in situ* avec une sonde antisens ABCA7.

La figure 7 représente le profil d'expression du gène codant pour la protéine ABCA7 sur une section de bronches d'une patiente asthmatique, par hybridation *in situ* avec une sonde antisens ABCA7.

10 La figure 8 représente le profil d'expression du gène codant pour la protéine ABCA7 sur une section de colon d'un patient atteint de la maladie de Crohn, par hybridation *in situ* avec une sonde antisens ABCA7.

15 La figure 9 représente le profil d'expression du gène codant pour la protéine ABCA7 sur une section de ganglion lymphatique, par hybridation *in situ* avec une sonde antisens ABCA7.

La figure 10 représente le profil d'expression du gène codant pour la protéine ABCA7 d'une section de synovie d'une patiente atteinte d'arthrite rhumatoïde, par hybridation *in situ* avec une sonde antisens ABCA7.

20 La figure 11 représente le profil d'expression du gène codant pour la protéine ABCA7 sur une section de peau d'une patiente atteinte de psoriasis par hybridation *in situ* avec une sonde antisens ABCA7.

EXEMPLES:

25 **EXEMPLE 1 : Détermination de l'extrémité 5' du cDNA de ABCA7**

Une amplification de l'extrémité du mRNA par RT-PCR (RACE) a été réalisée en utilisant le kit d'amplification SMART RACE cDNA (Clontech, Palo Alto, CA). Des ARNm (polyA) extraits à partir des tissus humains de la rate ont été utilisés comme 30 matrice pour produire une banque de cDNA 5' SMART selon les instructions du constructeur. Les amorces de première amplification et les amorces internes ont été choisies à partir de la séquence cDNA. Les amplifications réalisées avec les amorces internes d'amplification PCR ont été clonés. Des clones spécifiques ont été ensuite

amplifiés en utilisant des amorces dont les séquences sont respectivement (CAGGAAACAGCTATGAC) et (GCCAGTGTGATGGATAT) et séquencés sur les deux brins. Enfin les primers ABCA7 L1 GCGGAAAGCAGGTGTTTCAC (SEQ ID N° :11) et ABCA7L2 CGATGGCAGTGGCTTGG (SEQ ID N° :12) ont été utilisés pour 5 identifier l'extrémité du cDNA de ABCA7 humain.

EXEMPLE 2 : Analyse du promoteur des gènes ABCA7 humain et murin

Le site d'initiation de la transcription a été situé sur les promoteurs des gènes 10 humain et murin de ABCA7 en utilisant les trois logiciels suivants: TSSG et TSSW (Solovyev et al., *Ismb* (1997) 5, 294-302) et NNPP (Reese MG, et al., 1999). Une prévision des sites de fixation des facteurs de transcription humains et murins a été effectuée en utilisant le programme de recherche de motifs MatInspector (Quandt et al., *Nucl Acid Research* (1995) 23(23) 4878-84). Le calcul des scores pour chaque site de 15 fixation des facteurs de transcription est effectué en utilisant la formule suivante : $(Of - Tf)/(Tf)^{1/2}$, dans laquelle Of est la fréquence d'observation d'un motif et Tf est la fréquence calculée d'un motif consensus. Afin de séparer les motifs qui ne sont pas considérés comme pertinents, une première étape de filtration a été réalisé en ajustant 20 le score « similarité de la matrice » du programme MatInspector au dessus de 0,85 et le score « similarité du core » au dessus de 0,99. Enfin, une analyse comparative des promoteurs inter-espèces a été réalisée selon ce qui est décrit par Werner T (*Models for prediction and recognition of eukaryotic promoters, Mammalian Genome* (1999) 10: 168-175) afin de définir des modules de transcription comprenant des sites ayant un motif similaire et présents à la fois sur les séquences humaine et murine de la 25 séquence en amont du gène ABCA7.

EXEMPLE 3: Expression préférentielle des gènes ABCA7 humain et murin dans les tissus hématopoïétiques

Le profil d'expression des polynucléotides selon la présente invention est déterminé selon les protocoles d'analyse de northern blot et de transcription inverse couplée à la PCR décrits notamment par Sambrook et al (ref. CSH Sambrook, J.,

Fritsch, E.F., and Maniatis, T. (1989). " Molecular Cloning : A Laboratory Manual, " 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.).

Par exemple, dans le cas d'une analyse par transcription inverse, un couple d'amorces synthétisées à partir de chacun des ADNc complets des gènes humain et murin ABCA7 pour détecter les ADNc correspondants. Les séquences de ces amorces sont présentés dans le tableau 4.

La réaction de polymérase en chaîne (PCR) est réalisée sur des matrices d'ADNc correspondant à des ARNm polyA⁺ rétrotranscrits. La transcription inverse en ADNc est réalisée avec l'enzyme SUPERSCRIPT II (GibcoBRL, Life Technologies) 10 selon les conditions décrites par le fabricant.

La réaction de polymérase en chaîne est réalisée selon des conditions standard, dans 20 µl de mélange réactionnel avec 25 ng de la préparation d'ADNc. Le mélange réactionnel est composé de 400 µM de chacun des dNTP, de 2 unités de Thermus aquaticus (Taq) ADN polymérase (Ampli Taq Gold ; Perkin Elmer), de 0,5 µM 15 de chaque amorce, de 2,5 mM MgCl₂, et de tampon PCR. Trente cycles de PCR (dénaturation 30 s à 94 °C, hybridation de 30 s décomposé comme suit lors des 30 cycles : 64°C 2 cycles, 61°C 2 cycles, 58°C 2 cycles et 55°C 28 cycles et une élongation d'une minute par kilobase à 72°C) sont réalisés après une première étape 20 de dénaturation à 94°C durant 10 min dans un appareil thermocycler Perkin Elmer 9700. Les réactions de PCR sont visualisées sur gel d'agarose par électrophorèse. Les fragments d'ADNc obtenus peuvent être utilisés comme sondes pour une analyse par Northern blot et peuvent également être utilisés pour la détermination exacte de la séquence polynucléotidique.

Dans le cas d'une analyse par Northern Blot, une sonde d'ADNc produite 25 comme décrit ci-dessus est marquée au ³²P grâce au système de marquage d'ADN High Prime (Boehringer) selon les instructions indiquées par le fabricant. Après marquage, la sonde est purifiée sur une microcolonne de Sephadex G50 (Pharmacia) selon les instructions indiquées par le fabricant. La sonde marquée et purifiée est alors utilisée pour la détection de l'expression des ARNm dans différents tissus.

30 Le northern blot contenant des échantillons d'ARN de différents tissus humains (Multiple Tissue northern ou MTN ; références (Human II, 7759-1, Human 7760-1, and Human Fetal II 7756-1, Clontech) est hybridé avec la sonde spécifique marquée désignée de ABCA7 (2637-4881pb).

Le protocole suivi pour les hybridations et lavages peut être soit directement celui décrit par le fabricant (Manuel d'utilisation PT1200-1) soit une adaptation de ce protocole en utilisant les méthodes connues de l'homme de l'art et décrites par exemple dans F.Ausubel et al (*Currents Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience NY* (1989)). On pourra ainsi faire varier par exemple les températures de préhybridation et d'hybridation en présence de formamide.

Par exemple on pourra utiliser le protocole suivant :

10 **1- Compétition des membranes et PRE HYBRIDATION :**

- Mélanger : 40µl ADN sperme de saumon (10mg/ml)
 - + 40 µl ADN placentaire humain (10mg/ml)
- 15 - Dénaturer 5 mn à 96°C, puis plonger dans la glace le mélange.
- Oter le SSC 2X et verser 4 ml de mix formamide dans le tube d'hybridation contenant les membranes.
- 20 - Ajouter le mélange des deux ADN dénaturés.
- Incubation à 42°C pendant 5 à 6 heures, avec rotation.

2- Compétition de la sonde marquée :

25

- Ajouter à la sonde marquée et purifiée 10 à 50 µl ADN Cot I, selon la quantité de séquences répétées.
- Dénaturer 7 à 10 mn à 95°C.

30

- Incuber à 65°C pendant 2 à 5 heures.

3- HYBRIDATION :

- Oter le mélange de pré-hybridation.
- Mélanger 40 µl ADN sperme de saumon + 40 µl ADN placentaire humain ; dénaturer
- 5 5 mn à 96°C, puis plonger dans la glace.
- Ajouter dans le tube d'hybridation 4 ml de mélange formamide, le mélange des deux ADN et la sonde marquée/ADN Cot I dénaturée.
- 10 - Incuber 15 à 20 heures à 42°C, avec rotation.

4- Lavages :

- Un lavage à température ambiante dans du SSC 2X, pour rincer.

15

- 2 fois 5 minutes à température ambiante SSC 2X et SDS 0,1%.

- 2 fois 15 minutes SSC 0,1X et SDS 0,1% à 65°C.

Après hybridation et lavage, le blot est analysé après une nuit d'exposition au contact d'un écran au phosphore révélé à l'aide du Storm (Molecular Dynamics, Sunnyvale, CA).

Les résultats présentés à la figure 5 montrent que le gène ABCA7 de souris est exprimé dans les tissus adultes. Une quantité plus élevée d'ARNm ABCA7 murin est détectée dans les tissus hématopoïétiques tels que la rate et le thymus, ce qui est cohérent avec l'expression de ABCA7 qui a été observée dans les lignées myélomonocytaires et lymphocytaires. Aucune expression du gène ABCA7 n'a été détectée dans les lignées cellulaires fibroblastiques.

La figure 4 montre un pattern d'expression similaire du gène ABCA7 humain avec toutefois un fort signal d'hybridation dans le foie fœtal.

30

EXEMPLE 4 : Analyse du profil d'expression génique pour les dysfonctionnements dans le métabolisme des lipides, ou dans la signalisation de l'inflammation

La vérification de l'altération du niveau d'expression du gène ABCA7 peut-être déterminé par hybridation de ces séquences avec des sondes correspondant aux ARNm provenant de tissus hématopoïétiques de sujets atteints ou non, selon les 5 méthodes décrites ci-dessous :

1. Préparation des ARN totaux, des ARNm poly(A)⁺ et de sondes de cDNA

Les ARN totaux sont obtenus à partir de tissus hématopoïétiques de sujets 10 normaux ou bien atteints par la méthode à l'isothiocyanate de guanidine (Chomczynski et al., *Anal Biochem* (1987) 162:156-159). Les ARNm poly(A)⁺ sont obtenus par chromatographie d'affinité sur colonnes d'oligo(dT)-cellulose (Sambrook et al., (1989) *Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York*) et les cDNA utilisés comme sondes sont obtenus par RT-PCR 15 (DeRisi et al., *Science* (1997) 278: 680-686) avec des oligonucléotides marqués avec un produit fluorescent (Amersham Pharmacia Biotech ; CyDyeTM).

2. Hybrideation et détection des niveaux d'expressions

20 Les lames de verre contenant les séquences selon la présente invention correspondant au gène ABCA7 sont hybrideées avec les sondes nucléotidiques préparées à partir de l'ARN messager de la cellule à analyser. L'utilisation du système Amersham/molecular Dynamics (Avalanche MicroscannerTM) permet la quantification différentielle des expressions des produits de séquences sur le type cellulaire sain ou 25 affecté.

EXEMPLE 5 : Test destiné au criblage de molécules activant ou inhibant l'expression du gène ABCA7.

30 Le test de criblage permet d'identifier une substance susceptible de moduler l'activité de synthèse de la protéine ABCA7.

5.1 Construction des plasmides d'expression contenant un acide nucléique régulateur du gène ABCA7 humain.

La région de l'acide régulateur du gène ABCA7 humain allant du nucléotide en position -1111 jusqu'au nucléotide en position -1, par rapport au site d'initiation de la transcription, peut être amplifiée par la technique PCR à l'aide du couple d'amorces spécifiques de la région décrite ci-dessus à partir d'ADN génomique humain présent dans un vecteur BAC d'une collection de vecteurs BAC humains.

Le fragment d'ADN amplifié est digéré par l'endonucléase de restriction Sal 1, puis inséré dans le vecteur PXP1 décrit par Nordeen et al. (*Bio Techniques*, (1988) 6 : 10 454-457), au niveau du site de restriction Sal 1 de ce vecteur. L'insert a ensuite été séquencé.

5.2 Culture cellulaire et transfection

Des cellules de la lignée CHO ou HELA (ATCC, Rockville, MD, USA) ont été mises en culture dans le milieu E-MEM (Minimun Essential Medium with Earle's Salts) additionné de 10% (v/v) de sérum de veau foetal (BioWhittaker, Walkersville, MD). Approximativement $1,5 \times 10^5$ cellules ont été distribuées dans chacun des puits d'une plaque de culture de 12 puits (2,5 cm), et ont été cultivées jusqu'à environ 50-70% de confluence, puis co-transformées avec 1 µg du plasmide Sal-Lucif et 0,5 µg du vecteur témoin pBetagal (CloneTech Laboratories Inc., Palo Alto, CA, USA) en utilisant le nécessaire Superfectin Reagent Kit (QIAGEN Inc., Valencia, CA, USA). Deux heures après l'addition de l'ADN, le milieu de culture est éliminé et remplacé par du milieu complet AMEM (Minimum Esential Medium Eagle's Alpha Modification). Après une durée de vingt heures, les cellules sont placées dans du milieu frais de type DMEM (Dulbecco's Minimum Essential Medium) additionné de 2 µg/ml de glutamine, 100 unités/ml de streptomycine et de 0,1% de sérumalbumine bovine (BSA, Fraction V), en présence ou non de molécules à différentes concentrations.

Les cellules sont récupérées 16 heures après le dernier changement de milieu en utilisant une solution de lyse (Lysis Solution) provenant du nécessaire Tropix Luciferase Assay Kit (Tropix Inc., Bedford, MA, USA). Le lysat cellulaire est divisé en fractions aliquotes qui sont utilisées pour quantifier les protéines en utilisant le nécessaire MicroBCA Kit (Pierce, Rockford, IL, USA) ainsi que pour quantifier la production de luciférase et de bêta-galactosidase en utilisant respectivement les

nécessaires Tropix Luciferase Assay Kit et Galacto-Light Plus Kit. Les tests sont réalisés selon les recommandations du fabricant. Les molécules actives sur le promoteur ABCA7 sont ensuite sélectionnées en fonction du ratio « activité luciférase/activité beta-galactosidase.

5

EXEMPLE 6 : Experiences d'hybridation in situ.

Des échantillons de tissus concernés dans la paraffine ont été hybridés avec des sondes ARN complémentaires marquées radioactives. Plus précisément, un 10 fragment du gène ABCA7 correspondant à la séquence nucléotidique allant du nucléotide 594 au nucléotide 1055 de la séquence GenBank désignée NM-019112 a été sous clonée dans la plasmide pCRII (Invitrogen).

Des sondes ARN antisens marquées à l'Uridine triphosphate ³⁵S ont été ensuite générées avec les ARN polymerases SP6 et T7, puis hybridées aux différentes 15 sections tissulaires.

Les différentes sections de tissus ont été digérées avec la protéinase K et hybridées avec les sondes précédemment décrites à une concentration égale à environ 3,5x10⁷ dpm/ml pendant 18 heures à 65°C. Les lames ont été ensuite traitées avec de l'ARNase A et lavées dans du SSC, 0,1X à 70°C pendant 2 heures, et ont été 20 recouvertes d'une émulsion photographique Kodak NTB-2, exposées pendant 7 jours à 4°C, puis révélées en utilisant une solution Kodak D-19.

Elles ont été enfin colorées à l'hématoxyline et l'éosine (H&E) et les images ont été réalisées en utilisant une caméra photo digitale DVC couplée à un microscope Nikon.

25 La figure 6 est une coupe de l'artère d'un homme de 92 ans ayant subi une amputation au-dessous du genou, et présentant de l'artériosclérose ainsi qu'une inflammation aiguë. On observe un faible marquage spécifique des macrophages dans les thrombus et dans le site d'infiltration inflammatoire dans la tunique adventice.

La figure 7 qui est une section de bronches prélevées lors de l'autopsie d'une 30 femme asthmatique âgée de 63 ans, montre un faible marquage des lymphocytes et des macrophages dans l'infiltrat inflammatoire sous-muqueux.

La figure 8 est une section de colon prélevé au cours de l'opération d'une femme de 81 ans présentant un diagnostic clinique de la maladie de Crohn. On

observe un marquage des macrophages et d'une sous-famille de lymphocytes dans la lamina propria.

La figure 9 correspond à une coupe d'un ganglion lymphatique prélevé au cours de l'opération d'un homme âgé de 48 ans. Dans le centre germinatif réactif, les 5 cellules ganglionnaires sont faiblement marquées, et des macrophages isolés sont également marqués dans le ganglion lymphatique.

La figure 10 qui représente une coupe de la synoviale d'une femme de 25 ans présentant un diagnostic clinique d'arthrite rhumatoïde, montre un fort marquage des histiocytes sous synoviaux et des macrophages.

10 La figure 11 qui représente une section de peau obtenue suite à la biopsie d'une femme de 55 ans atteinte de psoriasis, montre un marquage moyen des macrophages dans l'infiltrat inflammatoire périvasculaire. Des lymphocytes isolés périvasculaires sont également marqués.

Tableau 1: Sites, scores, consensus, et positions par rapport au point d'initiation de la transcription (TSS) prédicts par les logiciels NNPP, TSSG et TSSW chez l'Homme

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-----------------------------------|-------------|------------------------|--------------------------|-----------|-------------------|-------------------|--------------------------|
| Analyse Comparative entre espèces | GFI1_01 | NNNNNNAAATCANNNNNNNNNN | gccacataAATCggagactolaga | 3.669729 | -569 | 1.00 | 0.98 |
| | HNF3B_01 | NNNTRTTTTRYTY | gaaTGTggcc | 3.978804 | -547 | 0.99 | 0.85 |
| | CEBPB_01 | RNRTRKINGMAAKNN | cgttgtGGAAaga | 1.857489 | -498 | 0.87 | 0.95 |
| | CEBPB_01 | RNRTRKINGMAAKNN | atcgtagGGAAcc | 1.857489 | -469 | 0.87 | 0.85 |
| | NF1_Q6 | NNTTGGCINNNNNNNNNNN | gctTGGccagccgggg | 1.651312 | -402 | 1.00 | 0.96 |
| | AP4_Q6 | CWCAGCTGGN | tgCAGGcggt | 12.133646 | -340 | 1.00 | 0.85 |
| | NFKAPPAB_01 | GGGAMTTYCC | GGAacccg | 9.285691 | -260 | 1.00 | 0.90 |
| | NFY_Q6 | TRRCCAATSRN | cgCCAAAtagc | 6.200634 | -106 | 1.00 | 0.89 |
| | AHRARNT_01 | KNNKNNTTYGGGTGCM | cgatggaggCGTGCtt | 10.450429 | -1065 | 1.00 | 0.87 |
| | CDPCR3HD_01 | NATYGATSSS | ggGATCeaagg | 2.474120 | -1004 | 1.00 | 0.87 |
| score Z >= 1.96 | IK1_01 | NNNTGGGAATRCC | caatggGAAaaatg | 14.464154 | -989 | 1.00 | 0.87 |
| | NFY_Q6 | TRRCCAATSRN | accaTTGGgg | 6.200634 | -978 | 1.00 | 0.87 |
| | LYF1_01 | TTGGGAGR | atGGGAGgg | 7.208594 | -975 | 1.00 | 0.91 |
| | BARBIE_01 | ATNNAAAGCNGRNG | accaAAAGctgaagc | 32.363969 | -963 | 1.00 | 0.91 |
| | E47_02 | NNNNMRCAGGTGTTNN | aggcaCAGGtgact | 15.631450 | -951 | 1.00 | 0.88 |
| | MYOD_01 | SRACAGGTGKYG | ccaaCAGGtgagt | 32.908282 | -949 | 1.00 | 0.89 |
| | LMO2COM_01 | SNNCAGGTGNNNN | ccaaCAGGtgagt | 2.232288 | -949 | 1.00 | 0.95 |
| | TH1E47_01 | NNNNGNRTRCTGGMWT | aggtagtCTGGtgagg | 16.677521 | -945 | 1.00 | 0.91 |
| | GFI1_01 | NNNNNNAAATCANNNNNNNNNN | ggggggaaatGATTggagg | 3.669729 | -934 | 1.00 | 0.92 |
| | NRF2_01 | ACGGGAAGNS | gggtTCCtggg | 6.109080 | -883 | 1.00 | 0.85 |
| CREL | NFKB_Q6 | NGGGAMTTCCNN | gagcgatTCCTC | 26.126380 | -861 | 1.00 | 0.88 |
| | NFKB_Q6 | NGGGAMTTCCNN | ggatTTTC | 2.943267 | -860 | 1.00 | 0.90 |
| | CREL_01 | SGGRNWTTCC | ggcgatTCCTC | 9.285691 | -859 | 1.00 | 0.91 |
| | NFKAPPAB_01 | GGGAMTTYCC | ggatTCCTCtcaa | 14.484154 | -858 | 1.00 | 0.89 |
| | IK1_01 | NNNTGGGAATRCC | TTCCCRKA | 6.281497 | -854 | 1.00 | 0.91 |
| | STAT_01 | TTCCCRKA | TTCCCTcaa | | | | |
| | BARBIE_01 | ATNNAAAGCNGRNG | ccatggCTTtgct | 32.363969 | -838 | 1.00 | 0.90 |
| USF | USF_Q6 | GYCACGTGNC | gtcGCGTggtc | 5.390268 | -802 | 1.00 | 0.94 |
| | AP2_Q6 | MKCCSCNGGG | tcCCCGtggg | 7.064136 | -777 | 1.00 | 0.86 |
| | VMYB_01 | AAYAACGGNN | ccCGTtgcc | 4.360548 | -776 | 1.00 | 0.89 |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|------------|---------------|----------------------|----------------------------|-----------|-------------------|--------------------|--------------------------|
| | TATA_01 | STATAAAWRNNNNNNNN | aactctataTTAacc | 7.166360 | -765 | 1.00 | 0.87 |
| | GATA_C | NGATAAGNMNN | ccatTTATCC | 2.004465 | -761 | 1.00 | 0.93 |
| | GATA1_03 | NNNNNGATAANNGNN | ctatTTATCctca | 2.776354 | -760 | 1.00 | 0.94 |
| | VMYB_01 | AAYAACGGNN | ccccAACGgca | 4.360548 | -743 | 1.00 | 0.91 |
| | AP2_Q6 | MKCCCSNCNGGGG | cggccGGGGgg | 7.064136 | -723 | 1.00 | 0.87 |
| | AHRARNT_01 | KNKNKNTYGCGTGCMNS | ccccCACGccctcact | 10.450429 | -707 | 1.00 | 0.85 |
| | BARBIE_01 | ATNNAAAGCNGRNNG | cgtAAAAGGcgigga | 32.363969 | -680 | 1.00 | 0.88 |
| | AP4_Q6 | CWCAGCTGGN | caaaGCTGtg | 12.132646 | -677 | 1.00 | 0.87 |
| | AHRARNT_01 | KNKNKNTYGCGGTGCMNS | cccaCACGGetcccaatt | 10.450429 | -664 | 1.00 | 0.87 |
| | HFH1_01 | NAWTGTTTATWT | aaagAGTTatTT | 51.812065 | -629 | 1.00 | 0.98 |
| | IK1_01 | NNNTGGGAATRCC | gggtGGGAaactgg | 14.484154 | -603 | 1.00 | 0.89 |
| | VMYB_01 | AAYAACGGNN | yyaAAGCGgt | 4.360548 | -598 | 1.00 | 0.88 |
| | CREL_01 | SGGRNWTTCC | cggttTTC | 6.194143 | -593 | 1.00 | 0.98 |
| | NFKAPPAB65_01 | GGGTTTCC | cggttTTC | 28.315415 | -593 | 1.00 | 0.95 |
| | GFI1_01 | NNNNNNAAATCANNNNNNNN | ttccatcaaAAATcagggtggccact | 3.669729 | -587 | 1.00 | 0.95 |
| | STAT_01 | TTCRKAA | TTCCTeaaa | 6.281497 | -587 | 1.00 | 0.88 |
| | STAT_01 | TTCRKAA | ttcgGGAA | 6.281497 | -496 | 1.00 | 0.92 |
| | BARBIE_01 | ATNNAAAGCNGRNNG | accctataCTTTacaa | 32.363969 | -459 | 1.00 | 0.85 |
| | TH1E47_01 | NNNNGNRTCTGGMWTT | agtcCAGGtttggaa | 16.677521 | -432 | 1.00 | 0.88 |
| | TH1E47_01 | NNNNGNRTCTGGMWTT | cccaaggAGCTGacta | 16.677521 | -429 | 1.00 | 0.90 |
| | AP4_Q6 | CWCAGCTGGN | gaCAGGCGgg | 12.132646 | -385 | 1.00 | 0.90 |
| | IK1_01 | NNNTGGGAATRCC | cagaGGAAacc | 14.484154 | -374 | 1.00 | 0.90 |
| | CHOP_01 | NNRTGCAATMCCC | tctTGCAtttcg | 22.328386 | -364 | 1.00 | 0.87 |
| | AP4_Q6 | CWCAGCTGGN | cggAGCTGcg | 24.429942 | -354 | 1.00 | 0.88 |
| | AP4_Q5 | NNCAAGCTGN | cggAGCTGcg | 3.791000 | -354 | 1.00 | 0.92 |
| | CHOP_01 | NNRTGCAATMCCC | cggtatTGCAGcc | 22.328386 | -346 | 1.00 | 0.93 |
| | HLF_01 | RTTACRYAAAT | GTAAcacaac | 12.981423 | -332 | 1.00 | 0.85 |
| | IK1_01 | NNNTGGGAATRCC | cgttgtCCGggag | 14.484154 | -285 | 1.00 | 0.88 |
| | SP1_Q6 | NGGGGGGGGGGYN | ggggGGGCGccig | 11.119144 | -276 | 1.00 | 0.86 |
| | NFKB_Q6 | NGGGGAMTTTCNN | cggGGAaccggccgg | 26.126380 | -262 | 1.00 | 0.87 |
| | AP4_Q6 | CWCAGCTGGN | cgCAGGCGccgg | 24.429942 | -146 | 1.00 | 0.88 |
| | AP4_Q5 | NNCAAGCTGN | cgCAGGCGccgg | 3.791000 | -146 | 1.00 | 0.92 |
| | ATF_01 | CNSTGACGTNNNYC | gagTGACggcgagg | 8.675151 | -121 | 1.00 | 0.86 |
| | AP1FLQ2 | RSTGACTNMNW | agtGAOgggca | 5.905504 | -120 | 1.00 | 0.91 |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|----------------|------|--------------------|----------------------|-----------|-------------------|-------------------|--------------------------|
| AP1_Q2 | | RSTGACTNMNW | agtGACGggca | 5.905504 | -120 | 1.00 | 0.89 |
| CAAT_01 | | NNNRRCCAAATSA | gtgcgcCCAAtag | 4.415584 | -108 | 1.00 | 0.86 |
| AHRARNT_01 | | KNIKKNNTYGGTGCMS | caatagcgCGTGCag | 10.450429 | -102 | 1.00 | 0.90 |
| AHRARNT_01 | | KNIKKNNTYGGTGCMS | aggaggggCGTGCcc | 10.450429 | -86 | 1.00 | 0.92 |
| GC_01 | | NRGGGGGGGGCNK | aaggGCCGggggc | 15.933816 | -28 | 1.00 | 0.92 |
| SP1_Q6 | | NGGGGGGGGGGYN | aaggGCCGggggc | 11.119144 | -28 | 1.00 | 0.93 |
| AP4_Q6 | | CWCAGCTGGN | gcctGCTGct | 12.133646 | -9 | 1.00 | 0.86 |
| SP1_Q6 | | NGGGGGGGGGYN | gcggGCGCggggg | 11.119144 | -2 | 1.00 | 0.90 |
| GC_01 | | NRGGGGGGGGCNK | gcggGGCGggggg | 15.933816 | -2 | 1.00 | 0.87 |
| IK1_01 | | NNNTGGAAATRCC | cggAGGGAGggcg | 14.484154 | 4 | 1.00 | 0.87 |
| AP4_Q6 | | CWCAGCTGGN | agaAGCTGcg | 12.133646 | 19 | 1.00 | 0.89 |
| IK1_01 | | NNNTGGAAATRCC | gagaGGGAaggag | 14.484154 | 56 | 1.00 | 0.87 |
| IK1_01 | | NNNTGGAAATRCC | caaggTCGGTggg | 14.484154 | 110 | 1.00 | 0.87 |
| IK1_01 | | NNNTGGAAATRCC | ccctGGGAalttg | 14.484154 | 116 | 1.00 | 0.92 |
| TST1_01 | | NNKGAWTTWANANTIN | tggAAATTaggggt | 6.882811 | 119 | 1.00 | 0.87 |
| NKX25_02 | | CWTTAATTG | gaATTAGg | 5.675005 | 122 | 1.00 | 0.91 |
| AP1_Q2 | | RSTGACTNMNW | tcTGACccctct | 5.905504 | 140 | 1.00 | 0.86 |
| AP1FJ_Q2 | | RSTGACTNMNW | tcTGACccctct | 5.905504 | 140 | 1.00 | 0.90 |
| RORA1_01 | | NWAWNNNAGGTCTAN | ctGACCCTccitc | 15.361241 | 141 | 1.00 | 0.94 |
| RORA2_01 | | NWAWNTAGGTCTAN | ctGACCCTccitc | 33.905118 | 141 | 1.00 | 0.85 |
| NRF2_01 | | ACCGGAAGNS | tccTTCCgggt | 6.109080 | 147 | 1.00 | 0.96 |
| ATF_01 | | CNSTGACGTNNNYC | tgttGACgcggct | 8.675151 | 160 | 1.00 | 0.91 |
| CREB_Q4 | | NSTGACGTMANN | gttGACgcggct | 5.543914 | 161 | 1.00 | 0.87 |
| AP1_Q2 | | RSTGACTNMNW | gttGACgcggct | 5.905504 | 161 | 1.00 | 0.89 |
| AP1FJ_Q2 | | RSTGACTNMNW | gttGACgcggct | 5.905504 | 161 | 1.00 | 0.90 |
| GF1_01 | | NNNNNNAAATCANNNNNN | gaatgtacactGATTcaagg | 3.669729 | 174 | 1.00 | 0.92 |
| CDFCR3HD_01 | | NATYGATSSS | aatttGATCac | 2.474120 | 175 | 1.00 | 0.97 |
| THE47_01 | | NNNNNGNRTCTGGWWTT | tccgacatCTGGggcc | 16.677521 | 203 | 1.00 | 0.89 |
| TALALPHA47_-01 | | NNNAACAGATGKTNNN | tccgacaTCTGGggcc | 43.162108 | 203 | 1.00 | 0.86 |
| TAL-BETA47_01 | | NNNAACAGATGKTNNN | tcggacatCTGGggcc | 43.162108 | 203 | 1.00 | 0.87 |
| E47_02 | | NNNNMRCAGGTGTTMNN | tcacacaCCTGcagcc | 15.631450 | 239 | 1.00 | 0.90 |
| E47_01 | | NSNGCAGGTGKNCNN | tcacacacCTGcagcc | 6.708124 | 239 | 1.00 | 0.97 |
| LMO2COM_01 | | SNNCAAGGTGNNN | acacaCCTGcag | 2.232288 | 241 | 1.00 | 0.97 |

| Filtration | Site | Consensus | Sequence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-----------------------------|----------------------|-------------------------|-------------------------|------------|-------------------|-------------------|--------------------------|
| | MYOD_01 | SRACAGGTGKYG | acacacCCTGag | 32.908282 | 241 | 1.00 | 0.87 |
| | VMYB_01 | AAYAACGGNN | ggccCGTTtaga | 4.360548 | 259 | 1.00 | 0.92 |
| | SRY_02 | NWWAACAAWANN | tcgttACAAatgg | 8.473458 | 293 | 1.00 | 0.86 |
| | TH1E47_01 | NNNNGNRRTCTGGMWNNT | tcctCCAGatcttag | 16.677521 | 320 | 1.00 | 0.88 |
| | E4BP4_01 | NRTTAYGTAAYN | cttgatGTTAAag | 12.6778534 | 341 | 1.00 | 0.86 |
| | VBP_01 | GTTACRTMAN | ttgtatGTTAAA | 5.053244 | 342 | 1.00 | 0.92 |
| | CREL_01 | SGGRNWTTCC | GGAAAGaaacc | 2.943267 | 352 | 1.00 | 0.85 |
| | VBP_01 | GTTACRTMAN | ctggccGTAAG | 5.053244 | 362 | 1.00 | 0.86 |
| | TH1E47_01 | NNNNGNRRTCTGGMWNNT | gtaaggagCTGGgict | 16.677521 | 367 | 1.00 | 0.92 |
| | AP1FJ_Q2 | RSTGACTNMNW | gcctggGTCAcc | 9.018855 | 398 | 1.00 | 0.92 |
| | AP1_Q2 | RSTGACTNMNW | gcctggGTCAcc | 9.018855 | 398 | 1.00 | 0.89 |
| | AP1_Q4 | RSTGACTMANN | gcctggGTCAcc | 13.1478586 | 398 | 1.00 | 0.86 |
| | ER_Q6 | NNARGNNNNNNNTGACCCYNN | ccctggGTCAcctttagcaac | 11.677290 | 399 | 1.00 | 0.88 |
| | ELK1_01 | NNNACMGGAAAGTNCNN | agaacatTCCGcccc | 15.164525 | 412 | 1.00 | 0.86 |
| | AP1FJ_Q2 | RSTGACTNMNW | ccctcgGTCAGc | 9.018855 | 426 | 1.00 | 0.93 |
| | AP1_Q2 | RSTGACTNMNW | ccctcgGTCAGc | 9.018855 | 426 | 1.00 | 0.92 |
| | AP1_Q4 | RSTGACTMANN | ccctcgGTCAGc | 13.1478586 | 426 | 1.00 | 0.89 |
| | GFI1_01 | NNNNNNAAATCANNNNNNNN | cgtcacggatcaGATTtcacatc | 3.669729 | 429 | 1.00 | 0.86 |
| | ATF_01 | CNSTGACGTNNNC | tcgtacgcGTCAGat | 8.675151 | 430 | 1.00 | 0.90 |
| | CREB_Q4 | NSTGACGTMANN | gttggccGTCAGa | 11.262690 | 431 | 1.00 | 0.88 |
| | CREB_Q2 | NSTGACGTAAAN | gttggccGTCAGa | 17.782892 | 431 | 1.00 | 0.86 |
| | AP1FJ_Q2 | RSTGACTNMNW | tcgtacgcGTCAGa | 5.905504 | 432 | 1.00 | 0.90 |
| | AP1_Q2 | RSTGACTNMNW | tcgtacgcGTCAGa | 5.905504 | 432 | 1.00 | 0.88 |
| TAL1BETAE47_0 | NNNAACAGATGKTNNN | tttcacatTCTGtgica | 64.766306 | 443 | 1.00 | 0.87 | |
| TAL1BETAITF2_0 ¹ | NNNAACAGATGKTNNN | tttcacatTCTGtgica | 64.766306 | 443 | 1.00 | 0.89 | |
| TAL1ALPHAE47_01 | NNNAACAGATGKTNNN | tttcacatTCTGtgica | 64.766306 | 443 | 1.00 | 0.89 | |
| AP1FJ_Q2 | RSTGACTNMNW | tcgtGTCAGa | 9.018855 | 450 | 1.00 | 0.93 | |
| AP1_Q4 | RSTGACTMANN | tcgtGTCAGa | 13.1478586 | 450 | 1.00 | 0.88 | |
| AP1_Q2 | RSTGACTNMNW | tcgtGTCAGa | 9.018855 | 450 | 1.00 | 0.91 | |
| CDPCR3HD_01 | NATYGATSSS | aaataGATCag | 2.474120 | 480 | 1.00 | 0.95 | |
| GFI1_01 | NNNNNNAAATCANNNNNNNN | agatcaggAAATCgcgtaccccg | 3.669729 | 483 | 1.00 | 0.89 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité dure | Similarité de la matrice |
|------------|------------|---------------------|-------------------------|------------|-------------------|-----------------|--------------------------|
| | AP1_Q2 | RSTGACTNNNN | 9ctGACcccg | 9.018855 | 495 | 1.00 | 0.94 |
| | AP1FJ_Q2 | RSTGACTNNNN | 9ctGACcccg | 9.018855 | 495 | 1.00 | 0.94 |
| | AP1_Q4 | RSTGACTMANN | 9ctGACTccgg | 13.148586 | 495 | 1.00 | 0.91 |
| | GATA1_03 | NNNNNGATAAANNNN | 9ctcTATCccggc | 2.776354 | 508 | 1.00 | 0.89 |
| | AP1FJ_Q2 | RSTGACTNNNN | ccTGACccgtt | 9.018855 | 529 | 1.00 | 0.92 |
| | AP1_Q2 | RSTGACTNNNN | ccTGACccgtt | 9.018855 | 529 | 1.00 | 0.91 |
| | BARBIE_01 | ATNNAAAGCNGRNGG | cctgacicTTTctct | 32.363969 | 529 | 1.00 | 0.86 |
| | AP1_Q4 | RSTGACTMANN | cctTGACccgtt | 13.148586 | 529 | 1.00 | 0.88 |
| | THE47_01 | NNNNNGNRTCTGGWWTT | ctttttCTGcccc | 16.677521 | 534 | 1.00 | 0.85 |
| | CP2_01 | GCNNINAMCMA | CTGccccgc | 3.245733 | 542 | 1.00 | 0.90 |
| | AP2_Q6 | MKCCCSNNGGG | cCCCCGgggtcc | 7.064136 | 546 | 1.00 | 0.98 |
| | GFI1_01 | NNNNNNAAATCANNNNNN | gccccctggcgGATTatgcgtac | 3.669729 | 554 | 1.00 | 0.85 |
| | AP4_Q6 | CWCAAGCTGGN | cagaGCTGgg | 12.133646 | 590 | 1.00 | 0.87 |
| | AP4_01 | WGARYCAGCTGYYGNNCNK | ggggcccaGCTGggccaa | 178.524329 | 603 | 1.00 | 0.86 |
| | RFX1_01 | NGNTNRCCNNRGGYACNN | cctcagccggGCAAcg | 7.226828 | 607 | 1.00 | 0.89 |
| | AP4_Q6 | CWCAAGCTGGN | tcaGCTGgg | 18.281794 | 608 | 1.00 | 0.97 |
| | AP4_Q5 | NNCAAGCTGNN | tcaAGCTGgg | 2.628244 | 608 | 1.00 | 0.96 |
| | AP4_Q6 | CWCAAGCTGGN | tccAGCGgg | 18.281794 | 608 | 1.00 | 0.93 |
| | AP4_Q5 | NNCAAGCTGNN | tcCAGCGgg | 2.628244 | 608 | 1.00 | 0.94 |
| | E47_01 | NSNGCAAGGTGKNNNN | cgggccaaCTGCctg | 6.708124 | 613 | 1.00 | 0.87 |
| | SREBP1_02 | KATCACCCAC | gtggggGTGATA | 30.499284 | 682 | 1.00 | 1.00 |
| | GATA1_03 | NNNNNGATAAANNNN | ggggGATAGccca | 2.776354 | 684 | 1.00 | 0.91 |
| | OLF1_01 | NNCNANTCCYNGRGAARNN | agccatTCCCctggcggtgg | 64.601270 | 697 | 1.00 | 0.89 |
| | IK1_01 | NNNTGGAAATRCC | gcaatTCCCcggt | 14.484754 | 698 | 1.00 | 0.87 |
| | NRF2_01 | ACGGGAAGNS | cacTTCCccgt | 6.109080 | 699 | 1.00 | 0.86 |
| | AHRARNT_01 | KNKNKNNTYGGTGCMS | tcctcgccGTTGiga | 10.450429 | 703 | 1.00 | 0.85 |
| | NFY_Q6 | TRRCCAAATSRN | cggCCAAatgtt | 6.200634 | 730 | 1.00 | 0.87 |
| | CDP_01 | CCATAATCGAT | ccAAATAtgtt | 147.430729 | 733 | 1.00 | 0.86 |
| | GATA_C | NGATAAGNNNN | tggttTATCt | 2.004465 | 744 | 1.00 | 0.93 |
| | GATA1_03 | NNNNNGATAAANNNN | gcgtttATCttgg | 2.776354 | 745 | 1.00 | 0.98 |
| | GFI1_01 | NNNNNNAAATCANNNNNN | ggggaaaggAAATCctggggct | 3.669729 | 770 | 1.00 | 0.89 |
| | CP2_01 | GCNNINAMCMA | CTGGggggggc | 3.245733 | 787 | 1.00 | 0.90 |
| | RORA1_01 | NWWNNNAGGTCA | gggtggGGTCAg | 7.616064 | 807 | 1.00 | 0.85 |

| Filtration | Site | Consensus | Sequence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la |
|------------|---------------------|---------------------------|------------|---------|-------------------|-------------------|------------------|
| | | | | | | | matrice |
| AP1_Q2 | RSTGACTNMNW | tggggGTCAgg | 5.905504 | 810 | 1.00 | 0.88 | |
| AP1FJ_Q2 | RSTGACTNMNW | 1ggggGTCAgg | 5.905504 | 810 | 1.00 | 0.91 | |
| NRF2_01 | ACCGGAAGNS | cctGGAAAGg | 6.109080 | 822 | 1.00 | 0.86 | |
| E47_02 | NNNMRCAGGTGTTNN | gciteCAGGigggct | 15.631450 | 832 | 1.00 | 0.86 | |
| ATF_01 | CNSTGAGCTTNNYC | tggTGAcaaaagg | 8.675151 | 860 | 1.00 | 0.92 | |
| CREB_Q4 | NSTGAGCTTMAIN | 9gTGAcaaaagg | 16.981467 | 861 | 1.00 | 0.94 | |
| AP1_Q2 | RSTGACTNMNW | 9gTGAcaaaagg | 9.018855 | 861 | 1.00 | 0.92 | |
| CREB_Q2 | NSTGACGTTAANN | 9gTGAcaaaagg | 26.730221 | 861 | 1.00 | 0.95 | |
| AP1FJ_Q2 | RSTGACTNMNW | 9gTGAcaaaagg | 9.018855 | 861 | 1.00 | 0.95 | |
| CREBP1_Q2 | NSTGAGCTTASN | 9gTGAcaaaagg | 22.127714 | 861 | 1.00 | 0.89 | |
| AP1_Q4 | RSTGACTTMANN | 9gTGAcaaaagg | 13.148586 | 861 | 1.00 | 0.91 | |
| CREB_01 | TGACGTMA | TGACGaaa | 4.176203 | 863 | 1.00 | 0.86 | |
| AP4_01 | WGARYCAGCTGGGNCK | aagicccaaGCTGicagg | 178.524329 | 917 | 1.00 | 0.88 | |
| AP4_Q6 | CWCAAGCTGGN | cacaGCTGic | 18.281794 | 922 | 1.00 | 0.97 | |
| AP4_Q6 | CWCAAGCTGGN | ccCAGCtGic | 18.281794 | 922 | 1.00 | 0.94 | |
| AP4_Q5 | NNCAAGCTGNN | cccaGCTGic | 2.628244 | 922 | 1.00 | 0.98 | |
| AP4_Q5 | NNCAAGCTGNN | ccCAGCtGic | 2.628244 | 922 | 1.00 | 0.96 | |
| AP1FJ_Q2 | RSTGACTNMNW | cagcGTCAgC | 5.905504 | 924 | 1.00 | 0.91 | |
| AP1_Q2 | RSTGACTNMNW | cagcGTCAgC | 5.905504 | 924 | 1.00 | 0.89 | |
| GFI1_01 | NNNNNNAAATCAGNNNNNN | tggcaacccAAATcagatggaggaa | 3.6589729 | 955 | 1.00 | 0.90 | |
| CAAT_01 | NNNNRCCAATSA | ggggGCAAAcA | 4.145584 | 956 | 1.00 | 0.98 | |
| NFY_C | NCTGATTGGTASY | ggggGCAAAcA | 69.836703 | 956 | 1.00 | 0.96 | |
| NFY_Q6 | TRRCAAATSRN | caggCCAAcA | 6.200634 | 958 | 1.00 | 0.96 | |
| AP4_Q6 | CWCAAGCTGGN | ggcggGCTGcg | 12.133646 | 976 | 1.00 | 0.86 | |
| AP2_Q6 | MRCCCSCNGGCG | cggggGGGGG | 7.064136 | 978 | 1.00 | 0.91 | |
| NFY_Q6 | TRRCAAATSRN | cocaTTGGGttt | 6.200634 | 995 | 1.00 | 0.95 | |
| CAAT_01 | NNNRCCAATSA | cocaTTGGGtttac | 4.415584 | 996 | 1.00 | 0.91 | |
| TATA_01 | STATAAAWRNNNNNN | ggggccatTTTAcgg | 7.166560 | 1025 | 1.00 | 0.86 | |
| GATA_C | NGATAAGNMNN | ccctttTATCgg | 2.004465 | 1029 | 1.00 | 0.92 | |
| GATA1_03 | NNNNNNGATAAANINGIN | cctttTATCgggg | 2.776354 | 1030 | 1.00 | 0.93 | |
| AP1_Q2 | RSTGACTNMNW | agTGACtactg | 9.018855 | 1040 | 1.00 | 0.93 | |
| AP1_Q4 | RSTGACTMANN | agtTGACtactg | 13.148586 | 1040 | 1.00 | 0.93 | |
| AP1FJ_Q2 | RSTGACTNMNW | agtTGACtactg | 9.018855 | 1040 | 1.00 | 0.94 | |
| GFI1_01 | NNNNNNAAATCAGNNNNNN | cicgcctAAATcagatggccagg | 3.669729 | 1056 | 1.00 | 0.94 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|-------------|--------------------|---------------------------|----------------------|------------|-------------------|--------------------|--------------------------|
| | | NNNN | | | | | |
| | STAT1_01 | NNNSANTTCCGGAAANTGN SN | cagagcttcAGGAAccatgc | 155.094175 | 1067 | 1.00 | 0.85 |
| | STAT1_01 | TTCCCRKAA | TTCCagggaa | 6.281497 | 1073 | 1.00 | 0.95 |
| | STAT1_01 | TTCCCRKAA | ttcaGGAA | 6.281497 | 1073 | 1.00 | 0.97 |
| GATA1_03 | NNNNGATAAANNGN | tggggGATAaaggaa | 2.776354 | 1090 | 1.00 | 0.95 | |
| GATA_C | NGATAAGNMNN | gGATAAaggaa | 2.004465 | 1094 | 1.00 | 0.94 | |
| BARBIE_01 | ATNNAAAGCNGRNNG | tcaGAAAGggggagg | 32.363968 | 1111 | 1.00 | 0.86 | |
| NFKB_Q6 | NGGGGAAATTCCNN | caGGGAGttccgg | 26.126380 | 1122 | 1.00 | 0.88 | |
| NFKBPAAB_01 | GGGAAATTYCC | GGAGgttgc | 9.285691 | 1124 | 1.00 | 0.93 | |
| AP2_Q6 | NIKCCSNCNGGG | tgCCGcgcgcg | 7.064136 | 1130 | 1.00 | 0.90 | |
| AP4_Q6 | CWCAGCTGNN | cgAGCCgcga | 12.133846 | 1134 | 1.00 | 0.86 | |
| XBP1_01 | NNNGTACGTGKNNNN | gcacgcACGtttcg | 21.302338 | 1141 | 1.00 | 0.85 | |
| VMYB_01 | AAYACGGNN | gacGGTTgtic | 4.360548 | 1161 | 1.00 | 0.93 | |
| ER_Q6 | NNARGNNNNNTGACCYNN | gaccgttgtctGACctct | 11.6777290 | 1161 | 1.00 | 0.86 | |
| AP1FJ_Q2 | RSTGACTNMNW | ctcGACcccttc | 2.752153 | 1170 | 1.00 | 0.90 | |
| ROF1_01 | NNWAWNNAGGTCAN | cGACCCttctgt | 7.616064 | 1171 | 1.00 | 0.92 | |
| NF1_Q6 | NNNTTGGCNNNNNNNNN | aaggaggggGCCAGaa | 1.651312 | -1093 | 1.00 | 0.94 | |
| DELTAEFL_01 | NNNCACTNNAN | agaAGGTggcc | 0.830664 | -1090 | 1.00 | 0.93 | |
| CMYB_01 | NNNNNNNGNCNGTTGN | ggccaaagaaGTTggct | 0.187475 | -1083 | 1.00 | 0.85 | |
| NF1_Q6 | NNNTTGGCNNNNNNNNN | agtGGCgtgtgggg | 1.651312 | -1074 | 1.00 | 0.91 | |
| MZF1_01 | NGNGGGGA | tgtGGGGA | -0.255601 | -1035 | 1.00 | 0.99 | |
| IK2_01 | NNNYGGGAWNNNN | tgtGGGGAGggaa | -1.019855 | -1035 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGGA | tggGGGGaa | -0.255601 | -1016 | 1.00 | 0.96 | |
| IK2_01 | NNNYGGGAWNNNN | tggGGGAGgggg | -1.019855 | -1016 | 1.00 | 0.89 | |
| MZF1_01 | NGNGGGGA | tggGGGGaa | -0.255601 | -1008 | 1.00 | 0.98 | |
| IK2_01 | NNNYGGGAWNNNN | tggGGGAataaa | -1.019855 | -1008 | 1.00 | 0.89 | |
| MZF1_01 | NGNGGGGA | caaGGGGaa | -0.255601 | -999 | 1.00 | 0.95 | |
| IK2_01 | NNNYGGGAWNNNN | ttttGGGAAat | -1.019855 | -999 | 1.00 | 0.91 | |
| MZF1_01 | NNNCACTNNAN | ggccGGCTcac | 0.830664 | -987 | 1.00 | 0.96 | |
| IK2_01 | NNNYGGGAWNNNN | ttttGGGAGgggg | -1.019855 | -978 | 1.00 | 0.93 | |
| AP4_Q5 | NNCAAGCTGNN | aaaggCTGaa | 0.302131 | -960 | 1.00 | 0.88 | |
| MYOD_Q6 | NNCANCTGNY | cacaggTGGag | 0.740149 | -948 | 1.00 | 0.91 | |
| DELTAEFL_01 | NNNCACTNNAN | cacAGGTggat | 0.830664 | -948 | 1.00 | 0.97 | |
| IK2_01 | NNNYGGGAWNNNN | ggccGGGAatg | -1.019855 | -913 | 1.00 | 0.90 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité d'acide aminé |
|------------|-------------------|-----------------------|-------------|-----------|-------------------|--------------------------|
| | IK2_01 | NNNNGGGAWNNNN | tagGGGAGagg | -1.019855 | -898 | 1.00 0.88 |
| NF1_Q6 | NNTTGGCNNNNNNCCNN | tggtctcgGCCAatt | 1.651312 | -885 | 1.00 0.86 | |
| IK2_01 | NNNNGGGAWNNNN | cagtTCCCTcaa | -1.019855 | -857 | 1.00 0.91 | |
| NF1_Q6 | NNTTGGCNNNNNNCCNN | cTTGGCtgacaciacc | 1.651312 | -831 | 1.00 0.93 | |
| DETAEF1_01 | NNNCACCTNAN | ctdACCTtag | 0.830664 | -820 | 1.00 0.88 | |
| NMYC_01 | NNNCACGTGNNNN | agtcICGTGccc | 0.303606 | -803 | 1.00 0.88 | |
| CMYB_01 | NNNNNNNGNCNTTGNN | algttcctcGTGgaga | 0.187475 | -782 | 1.00 0.90 | |
| IK2_01 | NNNNGGGAWNNNN | lgctTCCCGgtt | -1.019855 | -781 | 1.00 0.88 | |
| MZF1_01 | NGNGGGAA | tccccgtt | -0.225601 | -777 | 1.00 0.96 | |
| VMB_02 | NSYAACGGN | ccCGTTggc | 0.327098 | -775 | 1.00 0.96 | |
| NF1_Q6 | NNTTGGCNNNNNNCCNN | cgtTGGcaacttatt | 1.651312 | -773 | 1.00 0.94 | |
| GATA1_04 | NNCWGATARNNNNN | ctatTATCctta | 1.128924 | -760 | 1.00 0.92 | |
| GATA1_02 | NNNNNGATANKGN | ctatTATCctta | 1.132907 | -760 | 1.00 0.89 | |
| LMO2COM_02 | NMGATANSG | attTATCct | 0.679593 | -758 | 1.00 0.88 | |
| CMYB_01 | NNNNNNNGNCNTTGNN | gccttCAAACggcaatgcaca | 0.187475 | -745 | 1.00 0.94 | |
| VMB_02 | NSYAACGGN | ccAAACGccc | 0.327098 | -743 | 1.00 0.98 | |
| DETAEF1_01 | NNNCACCTNAN | tgccACCTtcct | 0.830664 | -732 | 1.00 0.93 | |
| IK2_01 | NNNNGGGAWNNNN | cggcGGGAgcgc | -1.019855 | -720 | 1.00 0.89 | |
| IK2_01 | NNNNGGGAWNNNN | gggtTCCCCcacg | -1.019855 | -712 | 1.00 0.91 | |
| MZF1_01 | NGNGGGAA | tCCCCCaag | 0.667940 | -708 | 1.00 0.99 | |
| IK2_01 | NNNNGGGAWNNNN | acttGCGccago | -1.019855 | -694 | 1.00 0.88 | |
| MZF1_01 | NGNGGGAA | tCCCGcgc | 0.667940 | -690 | 1.00 0.97 | |
| AP4_Q5 | NNCAGCTGNN | ccCAAGCgcct | 0.302731 | -688 | 1.00 0.86 | |
| AP4_Q5 | NNCAGCTGNN | caaaGCTGtg | 1.465487 | -677 | 1.00 0.90 | |
| IK2_01 | NNNNGGGAWNNNN | acgtTCCcaatt | -1.019855 | -660 | 1.00 0.91 | |
| AP4_Q5 | NNCAGCTGNN | tCAGCtca | 0.302731 | -650 | 1.00 0.87 | |
| DETAEF1_01 | NNNCACCTNAN | cttcACCTtca | 0.830664 | -645 | 1.00 0.95 | |
| IK2_01 | NNNNGGGAWNNNN | dagtGGAAaagg | -1.019855 | -603 | 1.00 0.96 | |
| VMB_02 | NSYAACGGN | ggtAACGgg | 0.327098 | -598 | 1.00 0.90 | |
| S8_01 | NNNNNNAAATTN | actatTAATcgaggact | -0.676808 | -566 | 1.00 0.86 | |
| NF1_Q6 | NNNTGGCNNNNNNCCNN | tgtTGGCccctccccc | 1.651312 | -544 | 1.00 0.95 | |
| IK2_01 | NNNNGGGAWNNNN | ccatTCCCcctc | -1.019855 | -537 | 1.00 0.87 | |
| MZF1_01 | NGNGGGAA | tCCCCCctc | -0.225601 | -533 | 1.00 0.96 | |
| IK2_01 | NNNNGGGAWNNNN | gtgtTCCCaagg | -1.019855 | -434 | 1.00 0.96 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-----------------------|----------------------------------|---|----------------------------|-----------|-------------------|-------------------|--------------------------|
| | IK2_01 | NNNYGGAWNNNN | tagGGGA gg ccct | -1.019855 | -410 | 1.00 | 0.88 |
| NF1_Q6 | NNTTGGC NNNNNN CCNNN | gaggaggccggGCCAgcc | 1.651312 | -408 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGAA | ccGGGGGaa | -0.225601 | -391 | 1.00 | 0.95 | |
| IK2_01 | NNNYGGGA WW NNNN | ccggGGGACaaGc | -1.019855 | -391 | 1.00 | 0.89 | |
| AP4_Q5 | NNCAGCTGNN | gacAGC GG gggg | 1.465487 | -385 | 1.00 | 0.92 | |
| IK2_01 | NNNYGGGA WW NNNN | agcgGGGA aa aga | -1.019855 | -382 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGAA | acGGGGGaa | -0.225601 | -382 | 1.00 | 0.99 | |
| IK2_01 | NNNYGGGA WW NNNN | cagaGGGA aa tc | -1.019855 | -374 | 1.00 | 0.94 | |
| CEPB _B _01 | RNRTKNNGMAAKNN | aaccttGCAAAtc | 1.857489 | -367 | 1.00 | 0.88 | |
| AP4_Q5 | NNCAGCTGNN | tgCAGGCGgggt | 1.465487 | -340 | 1.00 | 0.90 | |
| ARNT_01 | NNNNNCAGTG NNNNNN | tatacaCGTGGggag | 0.305357 | -330 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGAA | cgtGGGGaa | 0.667940 | -323 | 1.00 | 0.99 | |
| IK2_01 | NNNYGGGA WW NNNN | cgtggGGAGggca | -1.019855 | -323 | 1.00 | 0.88 | |
| IK2_01 | NNNYGGGA WW NNNN | tggtCTCC CC aaaa | -1.019855 | -308 | 1.00 | 0.89 | |
| MZF1_01 | NGNGGGAA | CCCCCaaa | -0.225601 | -304 | 1.00 | 0.96 | |
| AP4_Q5 | NNCAGCTGNN | gacAGC G ggag | 0.302731 | -286 | 1.00 | 0.86 | |
| IK2_01 | NNNYGGGA WW NNNN | tggtCTCCGggag | -1.019855 | -284 | 1.00 | 0.94 | |
| CETSP54_01 | NCMGGAWGYN | ccCGGA gg gc | 1.032772 | -279 | 1.00 | 0.89 | |
| IK2_01 | NNNYGGGA WW NNNN | gcgtGGGA cc tg | -1.019855 | -264 | 1.00 | 0.92 | |
| NF1_Q6 | NNTTGGC NNNNNN CCNNN | ccggggacttcGCCA cc cc | 1.851312 | -252 | 1.00 | 0.87 | |
| CEPB _B _01 | RNRTKNNGMAAKNN | tgatgtGCAA gg gg | 1.857489 | -229 | 1.00 | 0.90 | |
| IK2_01 | NNNYGGGA WW NNNN | ggcg T CCCGtta | -1.019855 | -211 | 1.00 | 0.91 | |
| IK2_01 | NNNYGGGA WW NNNN | gtgtATCCGtgtt | -1.019855 | -172 | 1.00 | 0.93 | |
| IK2_01 | NNNYGGGA WW NNNN | ctcc T CCCGtgg | -1.019855 | -162 | 1.00 | 0.88 | |
| NF1_Q6 | NNTTGGC NNNNNN CCNNN | cc T GGC cc gggggggtc | 1.851312 | -156 | 1.00 | 0.92 | |
| NF1_Q6 | NNTTGGC NNNNNN CCNNN | cgacggaggggGCCA gg | 1.651312 | -138 | 1.00 | 0.86 | |
| CREB_02 | NINGNTGACGYN | tgat T GA GG gc | 0.972541 | -122 | 1.00 | 0.88 | |
| AP4_Q5 | NNCAGCTGNN | ccggcGCTGct | 0.302731 | -70 | 1.00 | 0.86 | |
| DELTAEF1_01 | NNNCACCTNAN | tgctACCTG gg g | 0.830664 | -64 | 1.00 | 0.85 | |
| AP4_Q5 | NNCAGCTGNN | ctCAGC G ca | 0.302731 | -45 | 1.00 | 0.87 | |
| NK25_01 | TYAAGTG | cAC T gg | 1.905547 | -38 | 1.00 | 0.93 | |
| NF1_Q6 | NNNTTGGC NNNNNN CCNNN | acttGCGtaagggggg | 1.651312 | -37 | 1.00 | 0.92 | |
| IK2_01 | NNNYGGGA WW NNNN | ggcg T CCCGtgc | -1.019855 | -18 | 1.00 | 0.90 | |
| AP4_Q5 | NNCAGCTGNN | ggctGCTGct | 1.465487 | -9 | 1.00 | 0.92 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|------------------|------------------|-----------|---------|-------------------|-------------------|--------------------------|
| AP4_Q5 | NNCAGCTGNN | tgcGCTGgg | 0.302731 | -6 | 1.00 | 1.00 | 0.90 |
| IK2_01 | NNNYGGGAWNNNN | cggGGGAaggc | -1.019855 | 4 | 1.00 | 1.00 | 0.93 |
| AP4_Q5 | NNCAGCTGNN | aaGGCTGgg | 1.465487 | 18 | 1.00 | 1.00 | 0.93 |
| DELTAEF1_01 | NNNCACCTNN | ggAAGGTgata | 0.830684 | 37 | 1.00 | 1.00 | 0.97 |
| IK2_01 | NNNYGGGAWNNNN | gagaGGGAGggaa | -1.019855 | 56 | 1.00 | 1.00 | 0.93 |
| IK2_01 | NNNYGGGAWNNNN | ggccGGGAaggc | -1.019855 | 96 | 1.00 | 1.00 | 0.90 |
| IK2_01 | NNNYGGGAWNNNN | ggggGGGAigca | -1.019855 | 100 | 1.00 | 1.00 | 0.91 |
| IK2_01 | NNNYGGGAWNNNN | aatgtTCCCGggg | -1.019855 | 111 | 1.00 | 1.00 | 0.91 |
| SB_01 | NNNNNNYAATTN | ccctggaaATTAggg | -0.676808 | 116 | 1.00 | 1.00 | 0.93 |
| IK2_01 | NNNYGGGAWNNNN | ccctGGGAatata | -1.019855 | 116 | 1.00 | 1.00 | 0.96 |
| CET1P54_01 | NCMGGAWGYN | tcctTCCGgt | 1.032772 | 147 | 1.00 | 1.00 | 0.95 |
| CMYB_01 | NNNNNNGNCNGTTGNN | tcggggatGTTGacga | 0.187475 | 151 | 1.00 | 1.00 | 0.85 |
| CREB_02 | NINGNTGACGYN | atgtTCAcgaag | 0.972541 | 159 | 1.00 | 1.00 | 0.93 |
| AP4_Q5 | NNCAGCTGNN | gaogGCTGaa | 0.302731 | 167 | 1.00 | 1.00 | 0.89 |
| IK2_01 | NNNYGGGAWNNNN | atctGGGAacct | -1.019855 | 209 | 1.00 | 1.00 | 0.93 |
| DELTAEF1_01 | NNNCACCTNN | acacACCTgca | 0.830664 | 241 | 1.00 | 1.00 | 0.96 |
| MYOD_Q6 | NNNCACCTGNY | caaaCACCTgca | -0.175605 | 242 | 1.00 | 1.00 | 0.91 |
| VMYB_02 | NSYAACGGN | ccCGTTAgaa | 0.327098 | 260 | 1.00 | 1.00 | 0.98 |
| IK2_01 | NNNYGGGAWNNNN | gactTCCCccit | -1.019855 | 275 | 1.00 | 1.00 | 0.90 |
| MZF1_01 | NGNGGGGA | tCCCCctt | -0.225601 | 279 | 1.00 | 1.00 | 0.97 |
| IK2_01 | NNNYGGGAWNNNN | ccactTCCCccag | -1.019855 | 316 | 1.00 | 1.00 | 0.88 |
| MZF1_01 | NGNGGGGA | 1CCCCcag | -0.225601 | 320 | 1.00 | 1.00 | 0.96 |
| IK2_01 | NNNYGGGAWNNNN | taagTCCCccit | -1.019855 | 332 | 1.00 | 1.00 | 0.91 |
| IK2_01 | NNNYGGGAWNNNN | gaggTCCCcagtt | -1.019855 | 383 | 1.00 | 1.00 | 0.94 |
| CET1P54_01 | NCMGGAWGYN | cagtTCCCGgc | 1.032772 | 390 | 1.00 | 1.00 | 0.93 |
| DELTAEF1_01 | NNNCACCTNN | ggttcACCTTta | 0.830664 | 402 | 1.00 | 1.00 | 0.95 |
| CEBPB_01 | RNRTRKNGMAAKNN | acctttGCAAcit | 1.857489 | 406 | 1.00 | 1.00 | 0.93 |
| DELTAEF1_01 | NNNCACCTNN | tcggAGGTggcc | 0.830664 | 457 | 1.00 | 1.00 | 0.94 |
| GATA1_02 | NNNNNGATANKGNIN | gttcTATCccagc | 1.132907 | 508 | 1.00 | 1.00 | 0.92 |
| GATA1_04 | NNCWGATARNNN | gttcTATCcccg | 1.128924 | 508 | 1.00 | 1.00 | 0.90 |
| LMO2COM_02 | NMGATANSG | cctCTATccc | 0.679533 | 510 | 1.00 | 1.00 | 0.93 |
| IK2_01 | NNNYGGGAWNNNN | tcttaTCCCAGcc | -1.019855 | 511 | 1.00 | 1.00 | 0.95 |
| IK2_01 | NNNYGGGAWNNNN | tgtctTCCGcgccg | -1.019855 | 543 | 1.00 | 1.00 | 0.89 |
| IK2_01 | NNNYGGGAWNNNN | ggcgTCCCTCtcg | -1.019855 | 551 | 1.00 | 1.00 | 0.91 |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|------------|-------|---------------------|--------------------|------------|-------------------|--------------------|--------------------------|
| | S8_01 | NNNNNNYATTN | tgaggcgATTAAatgc | -0.6766808 | 559 | 1.00 | 0.86 |
| DETAEF1_01 | | NNNCACCTNAN | ataAGGTgtgg | 0.830664 | 578 | 1.00 | 0.97 |
| AP4_Q5 | | NNCAGCTGNIN | cagaGCTGgg | 1.465487 | 590 | 1.00 | 0.91 |
| CMYB_01 | | NNNNNNNGCNGTTGNIN | tgggCAAAGcgctgttc | 0.187475 | 614 | 1.00 | 0.84 |
| NIC25_01 | | TYAAGTG | caCTTct | 1.905547 | 670 | 1.00 | 0.88 |
| GATA1_02 | | NNNNNGATANKGNIN | gggggtGATAgcacca | 1.1329507 | 684 | 1.00 | 0.93 |
| GATA1_04 | | NNCWGATARNINN | gggtGATAgtccaa | 1.128924 | 685 | 1.00 | 0.92 |
| LMO2COM_02 | | NMGATANSG | gtGATAgtc | 0.679593 | 687 | 1.00 | 0.92 |
| IK2_01 | | NNNYGGGAWNNN | ccatTCOQcggg | -1.0198855 | 699 | 1.00 | 0.89 |
| NIC25_01 | | TYAAGTG | caCTTcc | 1.905547 | 699 | 1.00 | 0.88 |
| MZF1_01 | | NGNGGGGA | tCCCCCgg | -0.225601 | 703 | 1.00 | 0.95 |
| NF1_Q6 | | NNTTGCNNNNNNCCNNIN | tgccatgcactGCCAata | 1.651312 | 721 | 1.00 | 0.92 |
| GATA1_04 | | NNCWGATARNINN | gcgtgtATCTcg | 1.128924 | 745 | 1.00 | 0.95 |
| GATA1_02 | | NNNNNGATANKGNIN | gggtgtATCTcg | 1.132907 | 745 | 1.00 | 0.92 |
| LMO2COM_02 | | NMGATANSG | tgTATCTt | 0.679593 | 747 | 1.00 | 0.94 |
| MZF1_01 | | NGNGGGGA | tgAGGGGA | -0.225601 | 766 | 1.00 | 0.97 |
| IK2_01 | | NNNYGGGAWNNN | tggcGSGAaagg | -1.0198855 | 766 | 1.00 | 0.90 |
| NF1_Q6 | | NNTTGGCNNNNNNCCNNIN | ggcgtgggtggGCCAGgg | 1.651312 | 785 | 1.00 | 0.85 |
| LMO2COM_01 | | SNNCAGGTGNNN | ttcCAGGtgagg | 0.773414 | 834 | 1.00 | 0.94 |
| DETAEF1_01 | | NNNCACCTNAN | tccAGGTgtgg | 0.830664 | 835 | 1.00 | 0.98 |
| MYOD_Q6 | | NNCANCTGNY | tccAGGTGag | 0.740149 | 835 | 1.00 | 0.90 |
| CREB_02 | | NNGNTGACGYN | cggtTGAAGaaaa | 0.972541 | 859 | 1.00 | 0.84 |
| IK2_01 | | NNNYGGGAWNNN | tcggTCCCGgaa | -1.019855 | 886 | 1.00 | 0.89 |
| NIC25_01 | | TYAAGTG | ttAAAGTc | 1.905547 | 915 | 1.00 | 0.86 |
| IK2_01 | | NNNYGGGAWNNN | taagTCCCagc | -1.019855 | 916 | 1.00 | 0.90 |
| MZF1_01 | | NGNGGGGA | tCCCCCagc | 0.667940 | 920 | 1.00 | 0.97 |
| AP4_Q5 | | NNCAGCTGNN | gtCAGCCcttg | 0.302731 | 929 | 1.00 | 0.86 |
| NF1_Q6 | | NNTTGGCINNNNNCCNNIN | cgcgcggcaGCCAatc | 1.651312 | 949 | 1.00 | 0.92 |
| NF1_Q6 | | NNTTGGCINNNNNCCNNIN | tccTGGCagccataga | 1.651312 | 952 | 1.00 | 0.89 |
| AP4_Q5 | | NNCAGCTGNN | ggcgGCTGgg | 1.465487 | 976 | 1.00 | 0.91 |
| IK2_01 | | NNNYGGGAWNNN | ggcgTCCCatgg | -1.019855 | 990 | 1.00 | 0.84 |
| GATA1_04 | | NNCWGATARNINN | cctttATCTcggt | 1.128924 | 1030 | 1.00 | 0.92 |
| GATA1_02 | | NNNNNGATANKGNIN | cctttATCTcgagg | 1.132907 | 1030 | 1.00 | 0.92 |
| LMO2COM_02 | | NMGATANSG | cttTATCgga | 0.679593 | 1032 | 1.00 | 0.95 |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|-----------------|-------------------|-----------------|-----------|-------------------|-------------------|--------------------------|
| | S8_01 | NNNNNNYAAATTN | gtcTAATcagaggtt | -0.676808 | 1059 | 1.00 | 0.85 |
| AP4_Q5 | NNCAGCTGN | ctggGCTGg | 0.302731 | 1084 | 1.00 | 0.87 | |
| IK2_01 | NNYGGGAWNNNN | cggGGGAaaa | -1.019855 | 1089 | 1.00 | 0.95 | |
| GATA1_02 | NNNNINGATANKGN | tgggGATAaaagg | 1.132907 | 1090 | 1.00 | 0.93 | |
| GATA1_04 | NNCWGATAA | ggggGATAaaagg | 1.128924 | 1091 | 1.00 | 0.93 | |
| LMO2COM_02 | NMGATANS | ggGATAaaag | 0.879593 | 1093 | 1.00 | 0.93 | |
| CMYB_01 | NNNNNNNGNCNTGNN | ggggggggAGTTGccgg | 0.187475 | 1118 | 1.00 | 0.86 | |
| IK2_01 | NNYGGGAWNNNN | ggggGGAGgtg | -1.019855 | 1120 | 1.00 | 0.89 | |
| AP4_Q5 | NNCAGCTGN | cgCAGCCgca | 1.465487 | 1134 | 1.00 | 0.90 | |
| ARNT_01 | NNNNNCAGTGNNNNN | caccggCAGGtcgtcg | 0.305357 | 1142 | 1.00 | 0.86 | |
| CMYB_01 | NNNNINGNCNTGNN | cagccggacGTTGtcct | 0.187475 | 1155 | 1.00 | 0.93 | |
| VMYB_02 | NSYAACGGN | acCGTTgtc | 0.327098 | 1162 | 1.00 | 0.97 | |
| IK2_01 | NNYGGGAWNNNN | tctgtTCCCgtcc | -1.019855 | 1179 | 1.00 | 0.91 | |
| IK2_01 | NNYGGGAWNNNN | ccggTCCCCtgc | -1.019855 | 1184 | 1.00 | 0.87 | |
| MZF1_01 | NGNGGGGA | tCCGGCtg | -0.225601 | 1188 | 1.00 | 0.95 | |
| HNF3B_01 | NNNNTTTTTRYT | ctcTGTGTTgtac | 3.978804 | -1106 | 0.99 | 0.84 | |
| CDPCR3HD_01 | NATTGATSS | cgtcGATGag | 2.474120 | -1068 | 0.93 | 0.85 | |
| USF_Q6 | GYCACCGTGN | ggCCACAGggg | 5.3026268 | -950 | 0.86 | 0.87 | |
| E47_01 | NSNGCAGTGNKNCNN | ggcACAGgtgggt | 6.708124 | -950 | 0.83 | 0.86 | |
| USF_Q6 | GYCACCGTGN | cacaGGTGGag | 10.960075 | -948 | 0.82 | 0.87 | |
| USF_C | NCACGGTGN | acAGGGTGA | 0.301857 | -947 | 0.86 | 0.92 | |
| CETSP54_01 | NCMGGAWGYN | ggctTCCTgg | 1.032772 | -883 | 0.93 | 0.95 | |
| CAAT_01 | NNNRCCAATSA | cctggCCATtgg | 4.415584 | -878 | 0.86 | 0.86 | |
| CEBPB_01 | RNRTKNGMAAKNN | cggTTCCGtcataat | 1.857489 | -857 | 0.87 | 0.86 | |
| AP2_Q6 | MKCCCSNCNGGC | gccccCcatacg | 7.064136 | -843 | 0.98 | 0.86 | |
| USF_C | NCACGTGN | CTCTGTgg | 0.301857 | -801 | 0.81 | 0.86 | |
| CETSP54_01 | NCMGGAWGYN | ctcTGGAGtc | 1.032772 | -787 | 0.85 | 0.92 | |
| CETSP54_01 | NCMGGAWGYN | cacctTCCTgc | 1.032772 | -729 | 0.93 | 0.92 | |
| CETSP54_01 | NCMGGAWGYN | caactTCACgc | 1.032772 | -642 | 0.85 | 0.89 | |
| CETSP54_01 | NCMGGAWGYN | tctTCCAGa | 1.032772 | -611 | 0.85 | 0.90 | |
| CEBPB_01 | RNRTKNGMAAKNN | ggTTTCctaaaaa | 1.857489 | -591 | 0.99 | 0.88 | |
| CEBPB_01 | RNRTKNGMAAKNN | gttTTCCGtcataat | 1.857489 | -590 | 0.87 | 0.90 | |
| GC_01 | NRGGGGGGGGCNK | ggggccGTCGccct | 15.363816 | -540 | 0.88 | 0.91 | |
| SP1_Q6 | NGGGGGGGGGGYN | ggcccccTCCccct | 11.119144 | -539 | 0.84 | 0.93 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|---------------|----------------------|--------------------|------------|----------|-------------------|--------------------|--------------------------|
| CET1P54_01 | | NCMGGAWGYN | ccctCCCTgc | 1.032772 | -531 | 0.93 | 0.87 |
| AP2_Q6 | MKCCCSNCGGCG | agCCCCGggac | 7.064136 | -394 | 0.98 | 0.86 | |
| AP2_Q6 | MKCCCSNCGGCG | agccccGGGac | 7.064136 | -394 | 0.98 | 0.88 | |
| RFX1_02 | NINGTNRNCNNRGYAAACNN | cggggacggggGGAAact | 7.228454 | -380 | 0.88 | 0.90 | |
| NFKAPPAB65_01 | GGGRATTTCC | GGGAactct | 14.122479 | -370 | 0.83 | 0.89 | |
| CET1P54_01 | NCMGGAWGYN | gaatTCCTgc | 2.674616 | -368 | 0.93 | 0.85 | |
| VMYB_01 | AAYAACGGNN | gcGGTTata | 4.360548 | -336 | 0.81 | 0.86 | |
| TATA_01 | STATAAAWNNNNNN | ttaTACacgtgggg | 7.166360 | -331 | 0.80 | 0.86 | |
| LYT1_01 | TTTGGGAGR | ccccccaaa | 7.208594 | -305 | 0.82 | 0.85 | |
| CEBPB_01 | RNRTKNNGMAAKNN | cctttaaGAAAccc | 1.857489 | -205 | 0.99 | 0.89 | |
| PA05_C | NGTGGTCTC | tGTGATCcc | 4.266174 | -173 | 0.90 | 0.86 | |
| CAAT_01 | NNNRRCCAATSA | geaggCCAGiga | 4.415584 | -131 | 0.85 | 0.90 | |
| USF_Q6 | GYCACGTCNC | ggccAGTgcg | 5.390268 | -128 | 0.86 | 0.89 | |
| AP1_C | NTGASTCAN | gtGAGTGCAC | 1.751881 | -123 | 0.85 | 0.87 | |
| AP1_C | NTGASTCAN | gtGAGTGCAC | 1.751881 | -123 | 0.86 | 0.86 | |
| NFKAPPAB_01 | GGGAMTTTC | GGGGtgtgc | 9.285681 | -81 | 0.90 | 0.87 | |
| USF_Q6 | GYCACGTCNC | ggcACACTggc | 5.390268 | -40 | 0.86 | 0.89 | |
| USF_C | NCACGTGN | gcACCTTgg | 0.301857 | -39 | 0.84 | 0.91 | |
| CET1P54_01 | NCMGGAWGYN | cctGGAAagg | 1.032772 | 34 | 0.85 | 0.88 | |
| RFX1_01 | NINGTNRNCNNRGYAAACNN | aaggTCCctggaaata | 7.228282 | 111 | 0.88 | 0.89 | |
| CLOX_01 | NNATCGATTANYNN | tgaATTGacactga | 81.878596 | 173 | 0.87 | 0.89 | |
| CDP_02 | NWNATCGATTANYNN | tgaATTGacactga | 37.346724 | 173 | 0.85 | 0.89 | |
| LMO2COM_01 | SNNCAGGTGNNN | ggacatCTGgaga | 0.773414 | 205 | 0.82 | 0.90 | |
| MYOD_Q6 | NNCANCTGNY | gaCATCtgggg | -0.177595 | 206 | 0.92 | 0.89 | |
| USF_C | NCACGTGN | acACCTgc | 0.301857 | 243 | 0.86 | 0.92 | |
| AP2_Q6 | MKCCCSNCGGCG | agCCCCGtgtcc | 7.064136 | 251 | 0.98 | 0.88 | |
| CMYB_01 | NNNNNNNGNGTTGNN | cccccgtccGGTTgac | 0.187475 | 253 | 0.84 | 0.85 | |
| CET1P54_01 | NCMGGAWGYN | gaactCCtgc | 2.674616 | 267 | 0.93 | 0.85 | |
| AP2_Q6 | MKCCCSNCGGCG | ctCCCCCtgtcc | 7.064136 | 278 | 0.98 | 0.86 | |
| CEBPB_01 | RNRTKNNGMAAKNN | aaatggaaGAAActg | 1.857489 | 299 | 0.89 | 0.92 | |
| VMYB_01 | AAYAACGGNN | agaAACTgag | 4.360548 | 305 | 0.88 | 0.87 | |
| CEBPB_01 | RNRTKNNGMAAKNN | gcgtatGTAagg | 1.857489 | 340 | 0.93 | 0.89 | |
| CEBPB_01 | RNRTKNNGMAAKNN | alggaaaGGAAaga | 1.857489 | 345 | 0.87 | 0.86 | |
| CEBPB_01 | RNRTKNNGMAAKNN | ccctggcGTAAagg | 1.857489 | 360 | 0.93 | 0.88 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|-----------------|----------------|-----------|---------|-------------------|-------------------|--------------------------|
| CETSTP54_01 | NCMGGAWGYN | acccTCTCtg | 1.032772 | 415 | 0.93 | 0.96 | |
| LMO2COM_01 | SNNCAGGTGNNN | citccatCTCtgt | 0.773414 | 445 | 0.82 | 0.90 | |
| MYOD_Q6 | NNCAGACTGNY | tcCATCtggt | -0.175805 | 446 | 0.92 | 0.91 | |
| AP1_C | NTGASTCAN | cTGTGTCAG | 1.751681 | 451 | 0.86 | 0.86 | |
| CLOX_01 | NNTATCGATTANYNN | aaaATAGalccgaa | 81.979836 | 478 | 0.81 | 0.85 | |
| CDP_02 | NWNATCGATTANYNN | aaaATAGalccgaa | 37.346724 | 478 | 0.81 | 0.88 | |
| CETSTP54_01 | NCMGGAWGYN | taAGGAatcg | 1.032772 | 486 | 0.93 | 0.88 | |
| GATA_C | NGATAAGNMNN | agctctCTATCC | 2.004465 | 507 | 0.89 | 0.92 | |
| GC_01 | NRGGGGGGGGNNK | tggGGCAGacgtg | 15.933816 | 584 | 0.81 | 0.86 | |
| CETSTP54_01 | NCMGGAWGYN | tcctCTCAGC | 1.032772 | 604 | 0.85 | 0.89 | |
| LMO2COM_01 | SNNCAGGTGNNN | ctcAGGCTggc | 0.773414 | 607 | 0.88 | 0.94 | |
| LMO2COM_01 | SNNCAGGTGNNN | ctcAGGCTggc | 0.773414 | 607 | 0.88 | 0.93 | |
| MYOD_Q6 | NNCAGACTGNY | tcAGGCTggg | 1.656102 | 608 | 0.92 | 0.90 | |
| MYOD_Q6 | NNCAGACTGNY | tcAGGCTggg | 1.656102 | 608 | 0.92 | 0.90 | |
| LMO2COM_01 | NNCAGGTGNNN | ggggAACTGcc | 0.773414 | 615 | 0.80 | 0.91 | |
| VMYB_01 | AAAYAACGGNN | ggCAAACTgc | 4.360548 | 616 | 0.88 | 0.86 | |
| VMYB_02 | NSYACGGN | ggcAACTgc | 0.327098 | 616 | 0.82 | 0.89 | |
| MYOD_Q6 | NNCAGACTGNY | 9gCAACGcc | -0.175805 | 616 | 0.87 | 0.97 | |
| GATA_C | NGATAAGNMNN | tGATAGtcag | 2.004465 | 688 | 0.89 | 0.88 | |
| AP2_Q6 | MKCCCSNCGGG | tCCCCGtggcg | 7.064136 | 702 | 0.98 | 0.88 | |
| USF_Q6 | GYCACTGNC | ggcgTGTGaa | 5.390268 | 710 | 0.86 | 0.87 | |
| CHOP_01 | NNRTGCAATMCC | ggTGTGAAatgc | 22.328386 | 713 | 0.80 | 0.86 | |
| CETSTP54_01 | NCMGGAWGYN | aatgtTCCAGC | 1.032772 | 719 | 0.85 | 0.86 | |
| OCT1_02 | NINGAATATKCANNN | ggccaaATTGcg | 11.865447 | 732 | 0.98 | 0.91 | |
| CDPCR3_01 | CACCRATANNTATNG | CAATtttgttgtgt | 92.378088 | 734 | 0.97 | 0.86 | |
| VMYB_01 | AAYAACGGNN | tgCTGTTatc | 4.360548 | 744 | 0.82 | 0.89 | |
| ST1AT_01 | TTGCCRKAA | ttagAGAA | 6.281497 | 754 | 0.81 | 0.88 | |
| CETSTP54_01 | NCMGGAWGYN | ggAGGAGgt | 1.032772 | 801 | 0.93 | 0.90 | |
| AP2_Q6 | MKCCCSNCGGG | aggctggGGGtc | 7.064136 | 806 | 0.98 | 0.85 | |
| CETSTP54_01 | NCMGGAWGYN | tcAGGAcctg | 1.032772 | 816 | 0.93 | 0.87 | |
| CETSTP54_01 | NCMGGAWGYN | cctTGGAAatag | 1.032772 | 822 | 0.85 | 0.89 | |
| CETSTP54_01 | NCMGGAWGYN | ggctTCCAGg | 1.032772 | 831 | 0.85 | 0.92 | |
| USF_Q6 | GYCACTGNC | tcaGGTGag | 10.960075 | 835 | 0.82 | 0.86 | |
| USF_C | NCACGTGN | ccAGGGTGA | 0.301857 | 836 | 0.86 | 0.92 | |

| Filtration | Site | Consensus | Séquence | Z score | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|----------------|-----------------|--------------|-----------|-------------------|-------------------|--------------------------|
| | SP1_Q6 | NGGGGGGGGGYN | tggGGTggggcc | 11.119144 | 847 | 0.82 | 0.87 |
| GC_01 | NRGGGGGGGGCNK | tggGGTggagct | 15.933816 | 847 | 0.87 | 0.91 | |
| USF_Q6 | GYCACGTGNC | gcGGTGac | 5.390268 | 857 | 0.82 | 0.90 | |
| CEBPB_01 | RNRTKNNGMAAKNN | tggtagAAAagg | 1.857489 | 860 | 0.99 | 0.91 | |
| MYOD_01 | SRACAGGTGKYG | cocaaGCTGta | 32.908282 | 921 | 0.83 | 0.86 | |
| LMO2COM_01 | SNNCAGGTGNNNN | cocaaGCTGta | 2.232288 | 921 | 0.88 | 0.92 | |
| LMO2COM_01 | SNNCAGGTGNNNN | cccaAGCtgta | 0.773414 | 921 | 0.88 | 0.93 | |
| MYOD_Q6 | NNCANCTGNY | ccAGCGtgc | 1.656102 | 922 | 0.92 | 0.98 | |
| MYOD_Q6 | NNCANCTGNY | cccaAGCTGtc | 1.656102 | 922 | 0.92 | 0.89 | |
| LMO2COM_01 | SNNCAGGTGNNNN | aatCAGA!gcga | 0.773414 | 963 | 0.82 | 0.89 | |
| MYOD_Q6 | NNCANCTGNY | ataGATGcg | -0.1175805 | 964 | 0.92 | 0.94 | |
| CEBPB_01 | RNRTKNNGMAAKNN | ggTTAC!ccacc | 1.857489 | 1001 | 0.93 | 0.90 | |
| GC_01 | NRGGGGGGGGCNK | ttactCACCCctg | 15.933816 | 1004 | 0.87 | 0.87 | |
| SP1_Q6 | NGGGGGGGGGYN | factcCACCCctg | 11.119144 | 1005 | 0.82 | 0.85 | |
| USF_Q6 | GYCACGTGNC | atcgAGTGac | 5.390268 | 1036 | 0.96 | 0.88 | |
| HNF3B_01 | NNNTRTRTRY | tacTGT!Tgcct | 3.978804 | 1046 | 0.99 | 0.92 | |
| CETS1P54_01 | NCMGGAWGYN | agtttCAGg | 1.032772 | 1070 | 0.85 | 0.92 | |
| CETS1P54_01 | NCMGGAWGYN | ccAGGAaccc | 1.032772 | 1075 | 0.93 | 0.89 | |
| CEBPB_01 | RNRTKNNGMAAKNN | ggataaaGGAA!gta | 1.857489 | 1094 | 0.87 | 0.89 | |
| CEBPB_01 | RNRTKNNGMAAKNN | aggtaGAAAagg | 1.857489 | 1107 | 0.99 | 0.94 | |
| GC_01 | NRGGGGGGGGGCNK | aaggGGCAggagt | 15.933816 | 1116 | 0.81 | 0.85 | |
| NFKB_C | NGGGACTTTCCA | aGGGAGtggccc | 42.313772 | 1123 | 0.88 | 0.90 | |

Tableau 2 : Sites, scores, consensus, et positions par rapport au point d'initiation de la transcription (TSS) prédicts par les logiciels NNPP, TSSG et TSSW chez la souris

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|-----------------------------------|-------------|-----------------------|-------------------------|-----------|-------------------|--------------------|--------------------------|
| Analyse comparative entre espèces | GFI1_01 | NNNN | tgcctcacAAATCCaggcaacat | 2.393233 | -842 | 1.00 | 0.88 |
| | HNFB3B_01 | NONNNNAATCANNNGNNNNNN | aactTATTgtt | 2.929849 | -825 | 1.00 | 0.85 |
| | CEBPB_01 | NNNTRTTTTRYT | tgtttcGAATtg | 1.460836 | -787 | 0.99 | 0.94 |
| | CEBPB_01 | RNRTKNNNGMAAKNN | atgTTGCGttaaaatg | 1.460836 | -760 | 1.00 | 0.91 |
| | NF1_Q6 | NNITGGCENNNGMAAKNN | tctTGGCtgggggggg | 2.199282 | -688 | 1.00 | 0.88 |
| | AP4_Q6 | CWCAAGCTGGN | caCAcGcgttg | 14.114396 | -386 | 1.00 | 0.87 |
| | NFKAPPAB_01 | GGGAGCTGCC | gggAGCTGcc | 11.12 | -301 | 1.00 | 0.88 |
| | NFY_Q6 | TRRCCAAATSRN | ccCTCAAGgc | 5.187369 | -156 | 1.00 | 0.89 |
| | HFH2_01 | NAWTGTTTTRTTT | aaaaaaACAAaaa | 56.365713 | -1211 | 1.00 | 0.94 |
| | SRY_02 | NWWAAACAAWANN | aaaacAAAAaac | 7.986442 | -1208 | 1.00 | 0.94 |
| Score Z>= 1,96 | HFH2_01 | NAWTGTTTTRTTT | acaaaACAAaaa | 56.365713 | -1205 | 1.00 | 0.97 |
| | SRY_02 | NWWAAACAAWANN | aaaaACAAAAaca | 3.860390 | -1203 | 1.00 | 0.95 |
| | HFH2_01 | NAWTGTTTTRTTT | aaaaacAAACaaa | 28.165126 | -1200 | 1.00 | 0.98 |
| | SRY_02 | NWWAAACAAWANN | aaaaACAAAAac | 3.860390 | -1198 | 1.00 | 0.94 |
| | HFH2_01 | NAWTGTTTTRTTT | acaaaACAAaaa | 56.365713 | -1194 | 1.00 | 0.87 |
| | SRY_02 | NWWAAACAAWANN | aaaaACAAAAac | 7.986442 | -1192 | 1.00 | 0.94 |
| | HFH2_01 | NAWTGTTTTRTTT | aaaaaaACAAaaa | 28.165126 | -1188 | 1.00 | 0.91 |
| | HFH1_01 | NAWTGTTTATWT | acaaaACACaaa | 28.079407 | -1188 | 1.00 | 0.87 |
| | SRY_02 | NWWAAACAAWANN | aaaaAAACAAaaa | 3.860390 | -1186 | 1.00 | 0.98 |
| | TATA_01 | STATAAAWRNNNNNN | caATAAAaccctcg | 3.965816 | -1181 | 1.00 | 0.89 |
| Analyse comparative entre espèces | NF1_Q6 | NNITGGCENNNGNNNNNN | gttTGGCtgggtgggg | 2.199282 | -1140 | 1.00 | 0.93 |
| | CHOP_01 | NNRTGCAATMCC | augTGCAGggccct | 20.432681 | -1104 | 1.00 | 0.85 |
| | SRY_02 | NWWAAACAAWANN | cgtcACAAaaat | 3.860390 | -1093 | 1.00 | 0.85 |
| | LYF1_01 | TTTGGGAGR | ttGGGAGa | 7.842719 | -1082 | 1.00 | 0.90 |
| | E2F_02 | TTTSGCC | gcggaaaa | 3.546279 | -1071 | 1.00 | 0.91 |
| | GATA1_03 | NNNNNGATAAANIGN | tgtaGATAgtatcg | 2.031674 | -1042 | 1.00 | 0.88 |
| | CDPCR3HD_01 | NATYGATSSS | gataGATCgg | 2.349950 | -1037 | 1.00 | 0.97 |
| | NFE2_01 | TGCTGASTCAY | ggCTGAGcic | 21.956203 | -1009 | 1.00 | 0.87 |
| | CHOP_01 | NNRTGCAATMCC | atcTGCACaaaccc | 20.432681 | -969 | 1.00 | 0.86 |
| | NF1_Q6 | NNTTGGCENNNGNNNNNN | aaatcaatcggttGCCCAGgg | 2.199282 | -938 | 1.00 | 0.85 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|---------------|------------------------|--------------------------|--------------------------|------------|-------------------|-------------------|--------------------------|
| | SRY_02 | NWWAACAAWANN | tgctTTGtgaaa | 3.860390 | -897 | 1.00 | 0.85 |
| | HFH1_01 | NAWIGTTATWT | aaaaAACccgt | 28.079407 | -888 | 1.00 | 0.86 |
| | POLY_C | CATAAANCYYYYCTN | aaATAAAAccaggtttttt | 177.269419 | -888 | 1.00 | 0.88 |
| | ISRE_01 | CAGTTTCWCTTYYCC | caGTTTTttttcc | 384.173196 | -880 | 1.00 | 0.86 |
| TAL1BETA47_01 | NNNAACAGATGKNNN | ggagacatCTGgat | 16.998313 | -859 | 1.00 | 0.85 | |
| | NNNNNAAAATCANNNGNNNNNN | cattggagAAATCtgccatcgt | 2.393233 | -854 | 1.00 | 0.86 | |
| | SRY_02 | NWWAACAAWANN | gcctcAAATccaa | 3.860390 | -840 | 1.00 | 0.85 |
| | RFX1_01 | NINGTNRCCNNGYACNN | tacaatccaggGCAAActa | 7.172878 | -837 | 1.00 | 0.87 |
| | GFI1_01 | NNNNNNNAAAATCANNNGNNNNNN | caggaaactttGATTctatctt | 2.393233 | -830 | 1.00 | 0.89 |
| | GFI1_01 | NNNNNNNAAAATCANNNGNNNNNN | ttgttcAAATCtggatatgg | 2.393233 | -820 | 1.00 | 0.89 |
| | GATA1_03 | NNNNNGATAAANNGN | cttagGATATggg | 2.031644 | -809 | 1.00 | 0.90 |
| | NFY_Q6 | TRROCAATSRN | gataTTGGgt | 5.187369 | -804 | 1.00 | 0.88 |
| | GFI1_01 | NNNNNNNAAAATCANNNGNNNNNN | gggtgtccacccGATTctgaaatt | 2.393233 | -798 | 1.00 | 0.89 |
| | E47_02 | NNNNMRCAGGTGTtMNN | gctgcacaCTGttttttt | 15.432640 | -786 | 1.00 | 0.88 |
| | LMO2COM_01 | SINCAAGGTGNNNN | tgccaaCTCTatt | 3.041667 | -794 | 1.00 | 0.93 |
| | MYOD_01 | SRACAGGTGKYG | lgccaaCTCTatt | 40.688075 | -794 | 1.00 | 0.92 |
| | SRY_02 | NWWAACAAWANN | gaaaTTGtctag | 3.860390 | -780 | 1.00 | 0.87 |
| | TH1E47_01 | NNNNGNRRTCTGGMWTT | gtccaaatCTGgtgg | 16.434630 | -694 | 1.00 | 0.85 |
| | CP2_01 | GCMNNAMCMAG | CTGGtgggg | 3.137246 | -686 | 1.00 | 0.88 |
| | GATA1_03 | NNNNNGATAAANNGN | cacggGATAAaagg | 2.031644 | -685 | 1.00 | 0.88 |
| | SRY_02 | NWWAACAAWANN | ggatACAAtagac | 3.860390 | -661 | 1.00 | 0.85 |
| | NRF2_01 | ACCGGAAGNS | accTCCggcc | 6.701850 | -640 | 1.00 | 0.87 |
| | ER_Q6 | NNARGNNNNNTGACCYNN | aaatggccTGACctcc | 10.374054 | -573 | 1.00 | 0.89 |
| | AP1F1_Q2 | RSTGACTNNNN | tcTGACtccaa | 5.142253 | -564 | 1.00 | 0.90 |
| | AP1_Q2 | RSTGACTNNNN | tcTGACtccaa | 5.142253 | -564 | 1.00 | 0.86 |
| | RORAI_01 | NWWAACAAWANN | ctGACCCtccaaag | 5.437913 | -563 | 1.00 | 0.93 |
| | GATA1_03 | NNNNNGATAAANNGN | ccacaGATAggcca | 2.031644 | -558 | 1.00 | 0.87 |
| | OCT1_06 | CWNAWTKW/GATRYN | agatatgcATGCa | 8.438364 | -552 | 1.00 | 0.85 |
| | OCT_C | CTNATTGCAATAY | atagCAAATtaa | 70.265881 | -520 | 1.00 | 0.88 |
| | NKX25_02 | CWTAATTG | aaATTAAat | 3.983418 | -514 | 1.00 | 0.86 |
| | TST1_01 | NNKGAWTWAANNTNN | aaatAATTaatttta | 4.120840 | -513 | 1.00 | 0.87 |
| | NKX25_02 | CWTAATTG | atTTAATta | 3.983418 | -512 | 1.00 | 0.87 |
| | NKX25_02 | CWTAATTG | taATTAAaa | 3.983418 | -510 | 1.00 | 0.87 |
| | SRY_02 | NWWAACAAWANN | aaaaACAAagggt | 3.860390 | -499 | 1.00 | 0.93 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|------------|-------------|------------------------|---------------------------|-----------|-------------------|-------------------|--------------------------|
| | NF1_Q6 | NNTTGGCNNNNNNNNNN | tgtTGGCaaacggccita | 2.199282 | -481 | 1.00 | 0.85 |
| | AHRARNT_01 | KNKKNINTYGGTGCMS | ggcACGCGttttatc | 14.600483 | -476 | 1.00 | 0.86 |
| | GF1_01 | NNNNNAATACTAAGNNNNNNNN | aaacccttAAATCccggccatcagg | 2.393233 | -472 | 1.00 | 0.91 |
| | GF1_01 | NNNNNAATACTAAGNNNNNNNN | ggcttaaagggtGATTccaggcc | 2.393233 | -413 | 1.00 | 0.97 |
| | NQX25_02 | CWTAATTG | aaATTAAaa | 3.983418 | -393 | 1.00 | 0.86 |
| | LYF1_01 | TTGGGAGR | tggGGAGa | 7.842719 | -333 | 1.00 | 0.89 |
| | NF1_Q6 | NNTTGGCNNNNNNNNNN | tgtggggatggGCCAtt | 2.199282 | -305 | 1.00 | 0.88 |
| | NFKB_Q6 | NGGGGAMTTCCNN | tggGAAGtgccat | 30.067903 | -303 | 1.00 | 0.87 |
| | NFKAPPAB_01 | GGGATTYCC | GGAGggcc | 10.381187 | -301 | 1.00 | 0.88 |
| | ER_Q6 | NNARGNINANNNTGACYYNN | ggacttcaggGTGACccgt | 10.374054 | -279 | 1.00 | 0.86 |
| | E47_02 | NNNNMRCAGGTGTTMNN | acttcAGGtgacccg | 15.432640 | -277 | 1.00 | 0.90 |
| | LM02COM_01 | SNNCAGGTGNNN | tcaCAGGtgacc | 3.041667 | -275 | 1.00 | 0.94 |
| | MYOD_01 | SRACAGGTGKY | tcaCAGGtgacc | 40.698075 | -275 | 1.00 | 0.89 |
| | SREBP1_01 | NATCAGTGY | cacccGtGAcc | 15.355630 | -274 | 1.00 | 0.86 |
| | AP1_Q4 | RSTGACTMANN | ggTGA Cccgtt | 11.246163 | -270 | 1.00 | 0.86 |
| | AP1_Q2 | RSTGACTNNNN | ggTGA Cccgtt | 7.895015 | -270 | 1.00 | 0.91 |
| | AP1F1_Q2 | RSTGACTNNNN | ggTGA Cccgtt | 7.895015 | -270 | 1.00 | 0.93 |
| | VMYB_01 | AAYAACGGNN | accGGTTgc | 3.427439 | -266 | 1.00 | 0.93 |
| | NF1_Q8 | NNTTGGCNNNNNNNNNN | gtccccaaaggaaGCCAaacc | 2.199282 | -242 | 1.00 | 0.90 |
| | PADS_C | NGTGGTCTC | lGTG3Tcc | 5.230232 | -169 | 1.00 | 0.89 |
| | GC_01 | NRGGGGGGGGGCKNK | tggccCGGCClccct | 35.800511 | -167 | 1.00 | 0.87 |
| | SP1_Q6 | NGGGGGGGGGGYN | ggtccCGGCClccct | 25.529462 | -166 | 1.00 | 0.86 |
| | NF1_Q6 | NNTTGGCNCNNNNNNNNNN | caaTGGCaaaggccgcgt | 2.199282 | -152 | 1.00 | 0.85 |
| | E47_02 | NNNNMRCAGGTGTTMNN | agtagCAGGtgaaata | 15.432640 | -134 | 1.00 | 0.92 |
| | E47_01 | NSNGCAGGTGKNCNN | gtggCAGGtgccata | 9.748242 | -133 | 1.00 | 0.88 |
| | LM02COM_01 | SNNCAGGTGNNN | tagCAGGtgccaa | 3.041167 | -132 | 1.00 | 0.96 |
| | MYOD_01 | SRACAGGTGKY | tagCAGGtgccaa | 40.698075 | -132 | 1.00 | 0.86 |
| | CHOP_01 | NNRTGCAATMCCC | aggTGCaaatcc | 20.432681 | -128 | 1.00 | 0.96 |
| | CAAT_01 | NNNRRCGAATSA | atataCCAAatag | 3.434507 | -122 | 1.00 | 0.90 |
| | NFY_Q6 | TRRCCAATSRN | taCCAAatagt | 5.187369 | -120 | 1.00 | 0.92 |
| | GC_01 | NRGGGGGGGGGGCNK | aggGGCGggccgt | 35.806311 | -103 | 1.00 | 1.00 |
| | SP1_Q6 | NGGGGGGGGGYNN | aggggGCGGgggt | 25.529462 | -103 | 1.00 | 0.99 |
| | BARBIE_01 | ATNAAAAGCNGRNIGG | aggggAAAGgggggg | 28.462018 | 6 | 1.00 | 0.91 |
| | NQX25_01 | TYAAGTG | gaAAAGTg | 3.534570 | 9 | 1.00 | 0.88 |

| Filtration | Site | Consensus | Sequence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|------------|-----------------|----------------------|------------------------|------------|-------------------|-------------------|--------------------------|
| | VMYB_01 | AAYAACGGNN | caAAACGgtg | 3.427439 | 34 | 1.00 | 0.90 |
| | GF11_01 | NNNNNNAAATCANNNNNNNN | ggtgagaaaATCccgggggggt | 2.393233 | 40 | 1.00 | 0.90 |
| | NFKB_Q6 | NGGGAMTTCCNN | tgacaaaaATCCcg | 30.067803 | 42 | 1.00 | 0.86 |
| | ZD_01 | NGGCTCYATCACYC | gaaggggGAGCct | 41.225196 | 64 | 1.00 | 0.91 |
| | THE47_01 | NNNNGNR1CTGGMWTT | cggatatCTGGgtat | 16.434630 | 75 | 1.00 | 0.88 |
| | SREBP1_02 | KATACCCCCAC | atggggTGAgg | 27.710802 | 94 | 1.00 | 0.94 |
| | NFE2_01 | TGCTGASTCAY | ggCTGAgtcac | 21.950203 | 108 | 1.00 | 0.87 |
| | USF_Q6 | GYCACGTGNC | ggCACGgttc | 6.857788 | 114 | 1.00 | 0.87 |
| | IK1_01 | NNNTGGGGAAATRC | cacgtTCCCTgtat | 14.853568 | 116 | 1.00 | 0.88 |
| | GATA1_03 | NNNNNGATAANNN | tcgtGATAatgg | 2.031644 | 121 | 1.00 | 0.93 |
| | NKX25_02 | CWTAATTG | gtATAATt | 3.983418 | 126 | 1.00 | 0.86 |
| | E47_02 | NNNNMRCAGGTGTTNN | ggtcCAGGgtcac | 15.432640 | 136 | 1.00 | 0.87 |
| | LMO2COM_01 | SNNCAGGTGNNNN | ttcCAGGgtcac | 3.041567 | 138 | 1.00 | 0.96 |
| | MYOD_01 | SRACAGGTGKYG | tttcAGGgtcac | 40.698075 | 139 | 1.00 | 0.86 |
| | GATA1_03 | NNNNNGATAANNN | tttcTATCTtcac | 2.031644 | 164 | 1.00 | 0.95 |
| | NRF2_01 | ACGGAAAGNS | tctTTCGgg | 6.701650 | 171 | 1.00 | 0.91 |
| | STAF_02 | MNNTCGGKMATKCMNNNGNN | cttcggggggatCTGGGaaaa | 341.256924 | 173 | 1.00 | 0.86 |
| | IK1_01 | NNNTGGGAATRC | gtttGGGAAatatt | 14.853568 | 183 | 1.00 | 0.92 |
| | LYF1_01 | TTTGGGAGR | tgGGGAA | 7.842719 | 184 | 1.00 | 0.88 |
| | AP4_Q6 | CWCAGCTGGN | caCAGCggic | 14.114396 | 210 | 1.00 | 0.90 |
| | AP1_Q2 | RSTGACTNMNW | cagggtCAcic | 5.142253 | 212 | 1.00 | 0.88 |
| | AP1F1_Q2 | RSTGACTNMNW | cagggtCAcic | 5.142253 | 212 | 1.00 | 0.90 |
| | TAL1BETA47_01 | NNNAACAGATGKTNNN | ggggcatCTGgtcac | 48.119467 | 214 | 1.00 | 0.89 |
| | TAL1ALPHAE47_01 | NNNAACAGATGKTNNN | ggggatCTGgtcac | 48.119467 | 214 | 1.00 | 0.88 |
| | TAL1BETAUT2_01 | NNNAACAGATGKTNNN | ggggatCTGgtcac | 48.119467 | 214 | 1.00 | 0.88 |
| | THE47_01 | NNNNGNR1CTGGMWTT | ggggatCTGgtcac | 16.434630 | 214 | 1.00 | 0.86 |
| | AP1F1_Q2 | RSTGACTNMNW | atcgGTCAcc | 7.895015 | 220 | 1.00 | 0.93 |
| | AP1_Q4 | RSTGACTMANN | atcgGTCAcc | 11.246163 | 220 | 1.00 | 0.86 |
| | AP1_Q2 | RSTGACTNMNW | atcgGTCAcc | 7.895015 | 220 | 1.00 | 0.90 |
| | ER_Q6 | NNARGNNNNNTGACCCYNN | tcggGTCAccgtggggac | 10.374054 | 221 | 1.00 | 0.86 |
| | NF1_Q6 | NNTTGGCCTNNNNNNNN | gaggggacccGCCAacc | 2.199282 | 233 | 1.00 | 0.95 |
| | NKX25_01 | | cACTTic | 3.534570 | 287 | 1.00 | 0.88 |
| | AP1F1_Q2 | RSTGACTNMNW | ggggctGTCacc | 5.142253 | 281 | 1.00 | 0.91 |

| Filtration | Site | Consensus | Séquence | score Z | Position TSS (pb) | Similarité du core | Similarité de la matrice |
|--------------|----------------------|----------------------------|--------------|-----------|-------------------|--------------------|--------------------------|
| | AP1_Q2 | RSTGACTNNNNW | ggccCTGTCacc | 5.142253 | 281 | 1.00 | 0.88 |
| | SREBP1_02 | KATCACCCAC | tgtCACCCCC | 27.770802 | 205 | 1.00 | 0.87 |
| THE47_01 | NNNNNGIRTCGGMWWTT | ccccCCAGacitcaa | 16.434630 | 297 | 1.00 | 0.86 | |
| IRF1_01 | SMAAGYGAAACC | aaTTTCAcilit | 81.006772 | 311 | 1.00 | 0.87 | |
| IRF2_01 | GAAAGYGAAASY | aaTTTCAcilitat | 59.661305 | 311 | 1.00 | 0.85 | |
| NF1_Q6 | NNNTGGCINNNNNNNNNNN | gagggaaaggccGCCAAtt | 2.199282 | 341 | 1.00 | 0.93 | |
| NFY_Q8 | TRRCCAAATSRN | ccgCCAAAttcc | 5.187369 | 350 | 1.00 | 0.89 | |
| OCT1_Q6 | NNNNATGAAATTAN | ccaaTTTCAcdtgatgg | 11.430842 | 363 | 1.00 | 0.87 | |
| OCT1_06 | CWNAWTKWSATRYN | ccaaTTTCAATGta | 8.438364 | 353 | 1.00 | 0.92 | |
| OCT1_07 | TNTATGNTAAATT | AAATTtcaatgtta | 27.046261 | 355 | 1.00 | 0.88 | |
| CEBP_C | NGWNNTKNGYAAKNNAYA | aaaccttggCAATTtcc | 23.615437 | 374 | 1.00 | 0.86 | |
| NF1_Q6 | NNNTGGCINNNNNNNNNNN | ctttGGCaaatccctt | 2.199282 | 377 | 1.00 | 0.95 | |
| NFKAPPAB5_01 | GGGRATTTCC | ggccaaTTCC | 30.669184 | 381 | 1.00 | 0.86 | |
| CREL_01 | SGGRNWTTCC | ggccaaTTCC | 7.203414 | 381 | 1.00 | 0.86 | |
| NFKAPPAB_01 | GGGAMTTYCC | gcaatTTCCC | -10.361187 | 382 | 1.00 | 0.85 | |
| IK1_01 | NNNTGGAAATRC | caatTTGGGtctc | 14.853568 | 383 | 1.00 | 0.86 | |
| AP1_Q4 | RSTGACTMANN | tctctGTCAgc | 11.246163 | 392 | 1.00 | 0.90 | |
| AP1FJ_Q2 | RSTGACTNNMW | tctctGTCAgc | 7.8985015 | 392 | 1.00 | 0.95 | |
| AP1_Q2 | RSTGACTNNMW | tctctGTCAgc | 7.8985015 | 392 | 1.00 | 0.95 | |
| ISRE_01 | CAGTTTCWCCTTYYCC | caGTTTCCatcatacg | 384.173196 | 405 | 1.00 | 0.85 | |
| IK1_01 | NNNTGGAAATRC | cagttTCCTtacatc | 14.853568 | 405 | 1.00 | 0.87 | |
| GATA1_03 | NNNNNGATAAANNGN | ttccctTATCggat | 2.031644 | 409 | 1.00 | 0.91 | |
| GATA1_03 | NNNNNGATAAANNGN | atgggtATCgtgaa | 2.031644 | 415 | 1.00 | 0.88 | |
| NF1_Q6 | NNNTGGCINNNNNNNNN | tcgtggatcgatGCCACag | 2.199282 | 422 | 1.00 | 0.86 | |
| TATA_01 | STATAAAWRNNNNNN | aaaATAAAataacgaa | 3.9865815 | 458 | 1.00 | 0.85 | |
| GFI1_01 | NNNNNNAAATCANNNNNNNN | aaataacgaaATCaggaaatggcggt | 2.393233 | 464 | 1.00 | 0.92 | |
| VMYB_01 | AAYAACGGNN | aaAAACCCaaa | 3.427439 | 464 | 1.00 | 0.85 | |
| AHRARNT_01 | KNKNKNNTYGGTGCMS | caggaaatggGTGtc | 14.600483 | 475 | 1.00 | 0.93 | |
| AP1_Q4 | RSTGACTMANN | ccTGA Ctcic | 11.246163 | 503 | 1.00 | 0.89 | |
| AP1_Q2 | RSTGACTNNMW | ccTGA Ctcic | 7.8985015 | 503 | 1.00 | 0.90 | |
| AP1FJ_Q2 | RSTGACTNNMW | ccTGA Ctcic | 7.8985015 | 503 | 1.00 | 0.93 | |
| BARBIE_01 | ATNNAAAGCNGRNGG | taccctCTTtgc | 29.452018 | 532 | 1.00 | 0.86 | |
| AP1FJ_Q2 | RSTGACTNNMW | ttTGA Ctcgg | 7.8985015 | 541 | 1.00 | 0.90 | |
| AP1_Q2 | RSTGACTNNMW | ttTGA Ctcgg | 7.8985015 | 541 | 1.00 | 0.88 | |

| Filtration | Site | Consensus | Séquence | score Z | Position TSS (pb) | Similarité ducore | Similarité de la matrice |
|------------|------------|-------------------|------------------|-----------|-------------------|-------------------|--------------------------|
| | AP1_Q4 | RSTGACTMANN | tttGACTccgg | 11.246163 | 641 | 1.00 | 0.88 |
| | GC_01 | NRGGGGCGGGCNK | ggggGGCCGGccct | 35.805311 | 550 | 1.00 | 0.91 |
| | SP1_Q8 | NGGGGCGGGGYN | ggggGGCCGGcccc | 25.529462 | 550 | 1.00 | 0.93 |
| | TH1E47_01 | NNNNGNRTRCTGGMWTT | ctttttCTGGtttc | 16.434630 | 565 | 1.00 | 0.86 |
| | AHRARNT_01 | KNKNKNNTYGGTGOMS | ccctggggCGTGact | 14.600483 | 580 | 1.00 | 0.86 |
| | LYF1_01 | TTGGGAGR | cttGGGAGc | 7.842719 | 581 | 1.00 | 0.86 |
| | AP1IFJ_Q2 | RSTGACTNMNW | cgtTGACtGttgc | 7.895015 | 589 | 1.00 | 0.92 |
| | AP1_Q2 | RSTGACTNMNW | cgTGACtGttgc | 7.895015 | 589 | 1.00 | 0.89 |
| | AP1_Q4 | RSTGACTMANN | cgTGACtGttgc | 11.246163 | 589 | 1.00 | 0.89 |
| | IK1_01 | NNNTGGGAATRC | tcgtttCCtCtct | 14.852568 | 613 | 1.00 | 0.88 |
| | E47_01 | NSNGCAGGTGNNN | aaggccggCTGCaaa | 8.748242 | 638 | 1.00 | 0.89 |
| | AP4_Q8 | CWAGGCTGNN | gcCAGGtGtga | 21.241746 | 641 | 1.00 | 0.91 |
| | AP4_Q5 | NCAGGCTGNN | gcCAGGtGtga | 3.030778 | 641 | 1.00 | 0.94 |
| | AP4_Q6 | CWAGGCTGNN | gcgaGCTGca | 21.241746 | 641 | 1.00 | 0.94 |
| | AP4_Q5 | NCAGGCTGNN | gcgaGCTGca | 3.030778 | 641 | 1.00 | 0.95 |
| | OCT1_Q6 | NNNNATGCAAAATNN | ccatggAAATgac | 11.430342 | 642 | 1.00 | 0.88 |
| | AP1_Q2 | RSTGACTNMNW | aaATGACacaga | 7.895015 | 651 | 1.00 | 0.94 |
| | AP1IFJ_Q2 | RSTGACTNMNW | aaATGACacaga | 7.895015 | 651 | 1.00 | 0.94 |
| | AP1_Q4 | RSTGACTMANN | aaATGACacaga | 11.246163 | 651 | 1.00 | 0.91 |
| | E47_02 | NNNNMRCAAGGTGTMNN | ggggccctCTGggcg | 15.432640 | 677 | 1.00 | 0.88 |
| | LM02COM_01 | SNNCAGGTGNNN | ggccacCTGgggg | 3.041567 | 679 | 1.00 | 0.98 |
| | NYOD_01 | SRACAGGTGKY_G | ggccacCTGgggg | 40.698075 | 679 | 1.00 | 0.89 |
| | VMYB_01 | ATTAACGGNN | ggggACGggaa | 3.427439 | 690 | 1.00 | 0.91 |
| | TH1E47_01 | NNNNGNRTRCTGGMWTT | atccgggtCTGGatg | 16.434630 | 705 | 1.00 | 0.90 |
| | AP1IFJ_Q2 | RSTGACTNMNW | gtGTGACccgg | 5.142253 | 723 | 1.00 | 0.89 |
| | AP1_Q2 | RSTGACTNMNW | gtGTGACccgg | 5.142253 | 723 | 1.00 | 0.88 |
| | ZID_01 | NGGCTCYATCAAC | ggccgtgtGAGGCC | 41.225198 | 726 | 1.00 | 0.89 |
| | BARBIE_01 | ATNNAAAGCNGRNNG | ccccaaGTTTaaac | 29.452018 | 750 | 1.00 | 0.92 |
| | GC_01 | NRGGGGCGGGGCNKK | aaatccGGCCccct | 35.805311 | 767 | 1.00 | 0.97 |
| | SP1_Q8 | NGGGGGCGGGGGYN | aggccCccCccct | 25.529462 | 786 | 1.00 | 0.95 |
| | CREL_01 | SGGRNWTCC | aggccCTTCC | 3.467858 | 793 | 1.00 | 0.92 |
| | TH1E47_01 | NNNNGNRTRCTGGMWTT | tttCCAGaccccaagc | 16.434630 | 797 | 1.00 | 0.94 |
| | CDF_02 | NWNATCGATTANNN | gcctttatCGATAGc | 24.128980 | 811 | 1.00 | 0.91 |
| | CLOX_01 | NNTATCGATTANNN | gcctttatCGATAGc | 50.240668 | 811 | 1.00 | 0.90 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité dure | Similarité de la matrice |
|------------|-------------|------------------------|-------------------------|------------|-------------------|-----------------|--------------------------|
| | | | | | | | |
| | GATA1_03 | NNNNNGATAANNN | tcatcGATAGccct | 2.031644 | 815 | 1.00 | 0.68 |
| | NF1_Q6 | NNNTTGGCNNNNNNCCNN | lagcccttcAGGCCatc | 2.199282 | 822 | 1.00 | 0.93 |
| | NRF2_01 | ACCGGAAAGNS | ccCTTCAGcc | 6.701850 | 825 | 1.00 | 0.65 |
| | GFI1_01 | NNNNNNAAATCANNNNNNNNNN | ttccagccAAATcggctacggac | 2.393233 | 828 | 1.00 | 0.88 |
| | NFY_C | NCATGGTASY | tcggCCAAATcggc | 65.593286 | 829 | 1.00 | 0.96 |
| | CAAT_01 | NNNRCCAATSA | tcggCCAAAtca | 3.434507 | 829 | 1.00 | 0.99 |
| | NFY_Q6 | TRRCAATSRN | cggCCAAAtcg | 5.187369 | 831 | 1.00 | 0.96 |
| | AP4_Q6 | CWCAAGCTGGN | gaggGCTGcg | 14.114296 | 849 | 1.00 | 0.86 |
| | IK1_01 | NNNTGGAAATRC | cgggttCCGatgg | 14.853568 | 862 | 1.00 | 0.91 |
| | NFY_Q6 | TRRCAATSRN | cccatTTGGcica | 5.187369 | 868 | 1.00 | 0.95 |
| | CAAT_01 | NNNRCCAATSA | ccatTTGcGact | 3.434507 | 869 | 1.00 | 0.91 |
| | AP1FJ_Q2 | RSTGACTNMNW | cattggCTAct | 7.895015 | 870 | 1.00 | 0.94 |
| | AP1_Q4 | RSTGACTMANN | cattggCTAct | 11.246163 | 870 | 1.00 | 0.91 |
| | AP1_Q2 | RSTGACTNMNW | cattggCTAct | 7.895015 | 870 | 1.00 | 0.91 |
| | OLF1_01 | NNCNANTCCYNGGARNNNNGN | gtcactTCCTtagtgaaatctt | 77.917723 | 875 | 1.00 | 0.86 |
| | IK1_01 | NNNTGGAAATRC | tcactTCCTtagt | 14.853568 | 876 | 1.00 | 0.87 |
| | SRY_02 | NWWAACAAWAIN | tgctCTTGTtgc | 3.860390 | 906 | 1.00 | 0.89 |
| | GFI1_01 | NNNNNNAAATCANNNNNNNNNN | atcttggggggAGATTatgggg | 2.393233 | 921 | 1.00 | 0.88 |
| | TATA_01 | STATAAAWRNNNNNN | ggggggatTTAtg | 3.965815 | 927 | 1.00 | 0.85 |
| | AP2_Q6 | MKCCCSNNGGG | gaCCCCcagaca | 12.284970 | 978 | 1.00 | 0.86 |
| | TH1E47_01 | NNNNNGNRTCTGGMMWTT | atcttgtCTGcggcc | 16.434630 | 987 | 1.00 | 0.89 |
| | NF1_Q6 | NNNTGGCNNNNNNCCNN | attttgtggggGCCACatc | 2.199282 | 989 | 1.00 | 0.86 |
| | AP4_Q6 | CWCAAGCTGG | caCAGCtcac | 14.114296 | 1004 | 1.00 | 0.88 |
| | AP4_Q6 | CWCAAGCTGGN | clocGCTGtt | 14.114296 | 1040 | 1.00 | 0.85 |
| | TH1E47_01 | NNNNNGNRTCTGGMMWTT | cggCCAAAGggatca | 16.434630 | 1052 | 1.00 | 0.86 |
| | WMAF_01 | NNNNTGGTGAATCAGCANNNN | cggtccagaGTCAtatgg | 168.513881 | 1052 | 1.00 | 0.87 |
| | AP1_Q2 | RSTGACTNMNW | ccagaGTCAtc | 7.895015 | 1066 | 1.00 | 0.93 |
| | AP1_Q4 | RSTGACTMANN | ccagaGTCAtc | 11.246163 | 1056 | 1.00 | 0.90 |
| | AP1FJ_Q2 | RSTGACTNMNW | ccagaGTCAtc | 7.895015 | 1056 | 1.00 | 0.92 |
| | SOX5_01 | NNAAACATNN | aaaaCAATaa | 0.681190 | 1185 | 1.00 | 0.99 |
| | AP4_Q6 | NNCAGCTGNN | ggggGCTGtc | 0.503656 | -1122 | 1.00 | 0.85 |
| | DELTAEF1_01 | NNNNCACCTNN | gtcAGGtgcac | 0.538360 | -1107 | 1.00 | 0.96 |
| | IK2_01 | NNYGGGAWNN | gttGGGAGaa | -0.854442 | -1083 | 1.00 | 0.91 |
| | CWYB_01 | NNNNNGCNGTGNIN | aaacagacaaGTTGcaG | 0.594660 | -1065 | 1.00 | 0.92 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|---------------|-----------------|--------------|----------|-------------------|-------------------|--------------------------|
| | GATA1_02 | NNNNNGATANKGN | tgtaGATAgatg | 0.930257 | -1042 | 1.00 | 0.91 |
| | GATA1_04 | NNCVGATARNNN | gigaGATAgatg | 0.653180 | -1041 | 1.00 | 0.94 |
| LMO2COM_02 | NMGATANSG | gagATAgat | 0.569272 | -1039 | 1.00 | 0.91 | |
| IK2_01 | NNNYGGGAWNN | agtcTCCtca | -0.854442 | -1044 | 1.00 | 0.89 | |
| AP4_Q5 | NNCAGCTGN | accAGCtcc | 0.508566 | -994 | 1.00 | 0.96 | |
| IK2_01 | NNNYGGGAWNN | catcTCCtca | -0.854442 | -909 | 1.00 | 0.88 | |
| SOX5_01 | NNAACATNN | cclaAAATcc | 0.681190 | -839 | 1.00 | 0.86 | |
| GATA1_02 | NNNNNGATANKGN | cttagGATATggg | 0.930257 | -809 | 1.00 | 0.93 | |
| GATA1_04 | NNCVGATARNNN | tttgGATAtggg | 0.653180 | -808 | 1.00 | 0.89 | |
| LMO2COM_02 | NMGATANSG | agGATAtg | 0.569272 | -806 | 1.00 | 0.92 | |
| DELTAEF1_01 | NNNCACCTNN | tgccACCTgt | 0.638360 | -794 | 1.00 | 0.97 | |
| MYOD_Q6 | NNCANCCTGN | ggCACtCgt | 0.781061 | -753 | 1.00 | 0.95 | |
| SOX5_01 | NNAACATNN | aaATTGtcia | 0.681190 | -779 | 1.00 | 0.86 | |
| SOX5_01 | NNAACATNN | ggtaCAATtc | 0.681190 | -855 | 1.00 | 0.86 | |
| GATA1_02 | NNNNNGATANKGN | caagGATAcaag | 0.930257 | -865 | 1.00 | 0.89 | |
| GATA1_04 | NNCVGATARNNN | acaggGATAcaag | 0.653180 | -684 | 1.00 | 0.88 | |
| LMO2COM_02 | NMGATANSG | aggATACaa | 0.569272 | -662 | 1.00 | 0.88 | |
| CEPB_B1 | RNRTKNNGMaRNN | accGtGCAacc | 1.460836 | -651 | 1.00 | 0.94 | |
| DELTAEF1_01 | NNNCACCTNN | tccgACCTraa | 0.538360 | -636 | 1.00 | 0.87 | |
| CEPB_B1 | RNRTKNNGMaRNN | tctTGCctgagt | 1.460836 | -620 | 1.00 | 0.87 | |
| IK2_01 | NNNNGGGAWNN | gatgtCCcatat | -0.854442 | -611 | 1.00 | 0.95 | |
| DELTAEF1_01 | NNNCACCTNN | tcGACCTcca | 0.538360 | -564 | 1.00 | 0.85 | |
| GATA1_02 | NNNNNGATANKGN | coacaGATAgcca | 0.930257 | -556 | 1.00 | 0.93 | |
| GATA1_04 | NNCVGATARNNN | cacaGATAgcca | 0.653180 | -555 | 1.00 | 0.94 | |
| LMO2COM_02 | NMGATANSG | caGATAgc | 0.569272 | -553 | 1.00 | 0.96 | |
| S8_01 | NNNNNYAATTN | accctTAATaaat | -1.397287 | -526 | 1.00 | 0.86 | |
| SB_01 | NNNNNYAATTN | ataagcaATTaaat | -1.397287 | -520 | 1.00 | 0.95 | |
| S8_01 | NNNNNYAATTN | gcacatTAATTaaat | -1.397287 | -516 | 1.00 | 0.97 | |
| S8_01 | NNNNNYAATTN | aaatTAATaaat | -1.397287 | -514 | 1.00 | 0.99 | |
| IK2_01 | NNNYGGGAWNN | ttatTCCTCggca | -0.864442 | -466 | 1.00 | 0.95 | |
| AP4_Q5 | NNCAGCTGN | ccCAGCactc | 0.508566 | -461 | 1.00 | 0.85 | |
| AP4_Q5 | NNCAGCTGN | caCAGCgggt | 1.794672 | -386 | 1.00 | 0.93 | |
| S8_01 | NNNNNYAATTN | cctcaaaATTaaata | -1.397287 | -369 | 1.00 | 0.93 | |
| MZF1_01 | NGNGGGGA | cttGGGGA | 0.437162 | -334 | 1.00 | 0.96 | |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|------------|--------------|-------------------|--------------------|-----------|-------------------|-------------------|--------------------------|
| | IK2_01 | NNNYGGAWNNNN | atggGGAGggg | -0.854442 | -334 | 1.00 | 0.89 |
| | IK2_01 | NNNYGGAWNNNN | tgtGGAGctg | -0.854442 | -305 | 1.00 | 0.87 |
| | MZF1_01 | NGNGGGAA | tgtGGGGA | 0.437162 | -305 | 1.00 | 0.99 |
| | AP4_Q5 | NNICAAGCTGNN | ggggGCTGcc | 1.794672 | -301 | 1.00 | 0.91 |
| | MYOD_Q6 | NNICANCTGNY | cacaGTTGac | 0.781061 | -274 | 1.00 | 0.91 |
| | DELTAEFP1_01 | NNNCACCTNN | cacaAGTgcac | 0.5393860 | -274 | 1.00 | 0.95 |
| | CMYB_01 | NNNNNNNGCNGTTGNN | caggGacacGTTGtcac | 0.594660 | -272 | 1.00 | 0.90 |
| | VMYB_02 | NSYAAAGGGN | ccCGTTGtc | 0.465812 | -265 | 1.00 | 0.85 |
| | IK2_01 | NNNYGGAWNNNN | gttgtCCCcct | -0.854442 | -262 | 1.00 | 0.91 |
| | MZF1_01 | NGNGGGAA | tCCCCctc | 0.437162 | -258 | 1.00 | 0.96 |
| | IK2_01 | NNNYGGAWNNNN | cgtgtCCCagg | -0.854442 | -245 | 1.00 | 0.93 |
| | AP4_Q5 | NNCAGCTGNN | tgCAGGcgga | 0.508366 | -221 | 1.00 | 0.90 |
| | CMYB_01 | NNNNNNNGCNGTTGNN | caggaaatctGTTGtcac | 0.594660 | -216 | 1.00 | 0.89 |
| | IK2_01 | NNNYGGAWNNNN | gttgtTCCCTta | -0.854442 | -208 | 1.00 | 0.91 |
| | AP4_Q5 | NNCAGCTGNN | ggggGCTGtg | 0.508366 | -175 | 1.00 | 0.86 |
| | IK2_01 | NNNYGGAWNNNN | gttgtTCCGccct | -0.854442 | -168 | 1.00 | 0.92 |
| | DELTAEFP1_01 | NNNCACCTNN | agcaAGTtgccaa | 0.5393860 | -131 | 1.00 | 0.95 |
| | MYOD_Q6 | NNICANCTGNY | agcaAGTtccaa | -0.147777 | -131 | 1.00 | 0.96 |
| | GATA1_02 | NNNNNGATANKGNNN | tgcataTATCcaata | -0.046677 | -125 | 1.00 | 0.92 |
| | GATA1_04 | NNCAGGTATNNNN | tgaaaTATCcaat | 0.653180 | -125 | 1.00 | 0.87 |
| | LMO2COM_02 | NMGATANSG | caaTATCca | 0.569272 | -123 | 1.00 | 0.92 |
| | NIKX25_01 | TTAACATG | cACTTaa | 1.519193 | -33 | 1.00 | 0.98 |
| | IK2_01 | NNNYGGGGAWNNNN | ggggTCCCGccg | -0.854442 | -18 | 1.00 | 0.89 |
| | MZF1_01 | NGNGGGAA | tCCCGccg | 0.437162 | -15 | 1.00 | 0.96 |
| | VMYB_02 | NSYAAAGGGN | cggAACGcgat | 0.465812 | 34 | 1.00 | 0.92 |
| | IK2_01 | NNNYGGGGAWNNNN | aaaaTCCCGcgat | -0.854442 | 46 | 1.00 | 0.90 |
| | MZF1_01 | NGNGGGAA | tCCCGcgat | 1.879353 | 50 | 1.00 | 0.95 |
| | DELTAEFP1_01 | NNNCACCTNN | ggaaAGTggat | 0.5393860 | 63 | 1.00 | 0.95 |
| | IK2_01 | NNNYGGGGAWNNNN | tgtGGAGtgcgt | -0.854442 | 82 | 1.00 | 0.89 |
| | MZF1_01 | NGNGGGAA | tgGGGGGaa | 0.437162 | 82 | 1.00 | 0.96 |
| | AP4_Q5 | NNCAGCTGNN | gtggGCTGcg | 0.508366 | 105 | 1.00 | 0.86 |
| | ARNT_01 | NNNNNNCACTGNNNNNN | tggacACGttccctg | 0.511281 | 111 | 1.00 | 0.88 |
| | IK2_01 | NNNYGGGGAWNNNN | acgtTCCGcgat | -0.854442 | 117 | 1.00 | 0.92 |
| | GATA1_02 | NNNNNGATANKGNNN | tccctGATAttttg | 0.930257 | 121 | 1.00 | 0.91 |

| Filtration | Site | Consensus | Séquence | score Z | PositionTSS (pb) | Similarité dure | Similarité de la matrice |
|-------------|-----------------|------------------|----------------|-----------|------------------|-----------------|--------------------------|
| | GATA1_04 | NNCYGATARNNNN | ccctGATTAatly | 0.653180 | 122 | 1.00 | 0.95 |
| | SB_01 | NNNNNNYATTN | ctgatTAATggatt | -1.397287 | 124 | 1.00 | 0.95 |
| LMO2COM_02 | NMGATANSG | ctGATTAatt | 0.569212 | 124 | 1.00 | 0.91 | |
| GATA_C | NGATAAGNMNN | IGATAATlIgg | 1.411097 | 125 | 1.00 | 0.90 | |
| MYOD_Q6 | NNCANCTGNY | tcagaGTCcc | -0.147777 | 139 | 1.00 | 0.91 | |
| DELTAEF1_01 | NNNGACCTNAN | tcAGGTCct | 0.558360 | 139 | 1.00 | 0.95 | |
| IK2_01 | NNNYGGAWNNN | actTCCTge | -0.854442 | 150 | 1.00 | 0.98 | |
| GATA_C | NGATAAGNMNN | cttcTTATCC | 1.411097 | 163 | 1.00 | 0.97 | |
| GATA1_04 | NNCYGATARNNN | ttcttATCtc | 0.653180 | 164 | 1.00 | 0.94 | |
| GATA1_02 | NNNNNGATANCKNN | tcctTATCttcc | 0.830257 | 164 | 1.00 | 0.95 | |
| LMO2COM_02 | NMGATANSG | ccTATCtCt | 0.558272 | 166 | 1.00 | 0.98 | |
| CET1P54_01 | NCMGGAWGYN | tcctTCCTgg | 1.2441487 | 171 | 1.00 | 0.94 | |
| IK2_01 | NNNYGGAWNNN | tcggGGGagt | -0.854442 | 175 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGAA | tcGGGGA | 0.437162 | 175 | 1.00 | 0.95 | |
| IK2_01 | NNNYGGAWNNN | gtgtGGGaaaa | -0.854442 | 183 | 1.00 | 0.97 | |
| AP4_Q5 | NNICAGCTGNN | caCAGCggtc | 1.794672 | 210 | 1.00 | 0.92 | |
| DELTAEF1_01 | NNNCACCTNN | ggtcACCTcg | 0.538360 | 224 | 1.00 | 0.94 | |
| IK2_01 | NNNYGGAWNN | togaGGAAcc | -0.854442 | 231 | 1.00 | 0.90 | |
| CMYB_01 | NNNNNINGCNGTTGN | ctgcAAACttacaccc | 0.594660 | 242 | 1.00 | 0.85 | |
| DELTAEF1_01 | NNNCACCTNN | cttcACCTcua | 0.538360 | 250 | 1.00 | 0.94 | |
| IK2_01 | NNNYGGAWNN | agtgtCCCaatt | -0.854442 | 260 | 1.00 | 0.93 | |
| MZF1_01 | NGNGGGAA | ccccCacc | 0.437162 | 281 | 1.00 | 0.85 | |
| NKX25_01 | TYAAGTGT | caCTTta | 1.519193 | 316 | 1.00 | 0.94 | |
| IK2_01 | NNNYGGAWNNN | aaagtCCCGag | -0.854442 | 332 | 1.00 | 0.88 | |
| MZF1_01 | NGNGGGAA | tcCCTcgag | 1.679363 | 336 | 1.00 | 0.95 | |
| CEBPB_01 | RNRTRKNNGMAAKNN | aacttGCAAtt | 1.460836 | 375 | 1.00 | 0.98 | |
| IK2_01 | NNNYGGGAWNNN | aattTCCTtc | -0.854442 | 384 | 1.00 | 0.92 | |
| IK2_01 | NNNYGGGAWNNN | agttCCCTta | -0.854442 | 406 | 1.00 | 0.94 | |
| GATA1_04 | NNCYGATARNNNN | tcctTATCgtta | 0.653180 | 409 | 1.00 | 0.93 | |
| GATA1_02 | NNNINGATANCKNN | tcctTATCgtat | 0.830257 | 409 | 1.00 | 0.97 | |
| LMO2COM_02 | NMGATANSG | ccCTATCgg | 0.558272 | 411 | 1.00 | 0.99 | |
| GATA1_02 | NNNINGATANCKNN | atcgTATCatgaa | 0.930257 | 415 | 1.00 | 0.92 | |
| GATA1_04 | NNCYGATARNNNN | atcgTATCgaga | 0.653180 | 415 | 1.00 | 0.91 | |
| LMO2COM_02 | NMGATANSG | cggTATCcat | 0.558272 | 417 | 1.00 | 0.96 | |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|------|----------------|----------------|-----------|-------------------|-------------------|--------------------------|
| CETS1P54_01 | | NCMGGAWGYN | cagtTCCGgg | 1.244487 | 445 | 1.00 | 0.92 |
| MZF1_01 | | NGNGGGA | cgGGGGaa | 0.437162 | 451 | 1.00 | 0.98 |
| IK2_01 | | NNYGGGAWNIN | cggGGGaaata | -0.854442 | 451 | 1.00 | 0.91 |
| IK2_01 | | NNYGGGAWNIN | ccgtTCCGcgac | -0.854442 | 497 | 1.00 | 0.90 |
| CETS1P54_01 | | NCMGGAWGYN | lgactTCGga | 1.244487 | 543 | 1.00 | 0.87 |
| CETS1P54_01 | | NCMGGAWGYN | tccGGGaggc | 1.244487 | 547 | 1.00 | 0.90 |
| IK2_01 | | NNYGGGAWNIN | cctgGGGAGcgf | -0.854442 | 580 | 1.00 | 0.92 |
| IK2_01 | | NNYGGGAWNIN | cagtTCCGatct | -0.854442 | 614 | 1.00 | 0.94 |
| CETS1P54_01 | | NCMGGAWGYN | agactTCGgg | 1.244487 | 659 | 1.00 | 0.86 |
| DELTAEF1_01 | | NNCACCTNAN | ggccACTTggg | 0.538860 | 679 | 1.00 | 0.95 |
| MYOD_Q6 | | NNCACNTGNY | gcCACCTggg | 0.781061 | 680 | 1.00 | 0.91 |
| VMYB_02 | | NSYAAACGGN | ggggAACGga | 0.465612 | 690 | 1.00 | 0.93 |
| GATA1_02 | | NNNNINGATANKNN | tcataGATggct | 0.930257 | 815 | 1.00 | 0.91 |
| GATA1_04 | | NNCGATARNINN | calcGATAgcct | 0.653180 | 818 | 1.00 | 0.88 |
| LMO2COM_02 | | NNGATANSG | tcGATAgcc | 0.569272 | 818 | 1.00 | 0.95 |
| AP4_Q5 | | NCAGCTGNN | atcAGGtacg | 0.508566 | 837 | 1.00 | 0.89 |
| AP4_Q5 | | NCAGCTGNN | ggcgGCTTgg | 1.794672 | 849 | 1.00 | 0.91 |
| IK2_01 | | NNNYGGGAWNIN | gggtTCGCatg | -0.854442 | 853 | 1.00 | 0.97 |
| IK2_01 | | NNNYGGGAWNIN | cactTCGClagt | -0.854442 | 877 | 1.00 | 0.92 |
| NKX25_01 | | TYAAAGTG | CACTTcc | 1.519193 | 877 | 1.00 | 0.88 |
| IK2_01 | | NNNYGGGAWNIN | tgcGGGAGagt | -0.854442 | 925 | 1.00 | 0.89 |
| AP4_Q5 | | NNCAGCTGNN | cICAGCCcga | 0.508566 | 945 | 1.00 | 0.87 |
| AP4_Q5 | | NNCAGCTGNN | caCAGCtcac | 1.794672 | 1004 | 1.00 | 0.92 |
| AP4_Q5 | | NNCAGCTGNN | ctcGCTGtt | 1.794672 | 1040 | 1.00 | 0.91 |
| CETS1P54_01 | | NCMGGAWGYN | tggtTCGcg | 1.244487 | 1046 | 1.00 | 0.85 |
| HFH2_01 | | NAWTGTTT | aaacaaAAAaaa | 56.365713 | -1219 | 0.82 | 0.89 |
| HNF3B_01 | | NNNTRTRTRY | aaacaaAAAaaa | 6.168471 | -1219 | 0.85 | 0.90 |
| HFH2_01 | | NAWTGTTT | aaaaaaAAAaaa | 28.165126 | -1215 | 0.82 | 0.88 |
| HNF3B_01 | | NNNTRTRTRY | aaaaaaAACaaa | 6.168471 | -1211 | 0.99 | 0.88 |
| HNF3B_01 | | NNNTRTRTRY | aaaaaaAAACaa | 9.407093 | -1207 | 0.85 | 0.89 |
| HNF3B_01 | | NNNTRTRTRY | aaaaaaAAACaa | 9.407093 | -1196 | 0.85 | 0.89 |
| HNF3B_01 | | NNNTRTRTRY | aaaaaaAAACaa | 9.407093 | -1190 | 0.85 | 0.89 |
| HFH2_01 | | NAWTGTTT | aaaaacAAATaaaa | 28.165126 | -1185 | 0.90 | 0.85 |
| TATA_C | | NCTAAAAAR | acAAATAAAAa | 8.111772 | -1182 | 0.89 | 0.93 |

autone
(Paramètres par défaut
MatInspector)

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité duree | Similarité de la matrice |
|------------|-------------|---------------------|------------------|-----------|-------------------|------------------|--------------------------|
| | VMYB_01 | AAYAACGGNN | tttCTGTTtct | 3.427439 | -1171 | 0.82 | 0.85 |
| | VMYB_01 | AAYAACGGNN | agaACAGac | 3.427439 | -1068 | 0.82 | 0.86 |
| | LM02COM_01 | SNICAGGTGNNN | acaCA GTgaaat | 1.242813 | -1060 | 0.80 | 0.87 |
| | MYOD_Q6 | NCANCTGNY | cacaGTGaa | -0.147777 | -1059 | 0.87 | 0.89 |
| | OCT1_06 | CWINAWTKWSATRYN | cacaGTgATGAA | 8.436364 | -1059 | 0.83 | 0.86 |
| | VMYB_02 | NSYAACGGN | atAGT Tgaa | 0.465812 | -1058 | 0.82 | 0.88 |
| | GATA_C | NGATAAGNMNN | aGATA GaaGcg | 1.411097 | -1038 | 0.89 | 0.90 |
| | CEBPB_01 | RNRTKNGMAAKNN | ggggggGAAAGag | 1.460336 | -1022 | 0.99 | 0.93 |
| | LM02COM_01 | SNICAGGTGNNN | atgcatCTGcaca | 1.242813 | -973 | 0.82 | 0.91 |
| | MYOD_Q6 | NCANCTGNY | tgCATCtgca | -0.147777 | -972 | 0.92 | 0.91 |
| | AP1_C | NTGASTCAN | cTAACtCac | 1.430304 | -940 | 0.86 | 0.87 |
| | AP1_C | NTGASTCAN | cTAACtCac | 1.430304 | -940 | 0.85 | 0.87 |
| | PADS_C | NGTGGTCTC | gGTGA Ttca | 5.230232 | -922 | 0.90 | 0.89 |
| | CEBP_C | NGWNTKNGKCYAAKNNAYA | tgcttggAAATaaacc | 23.615437 | -897 | 0.89 | 0.89 |
| | CEBPB_01 | RNRTKNGMAGAKNN | gcttggGAAAtaa | 1.460336 | -896 | 0.99 | 0.95 |
| | VMYB_01 | AAYAACGGNN | acAA GTTTttt | 3.427439 | -882 | 0.88 | 0.88 |
| | HFH2_01 | NAWGTGTTTTT | cagTTTttttttt | 28.185126 | -880 | 0.82 | 0.86 |
| | CETS1P54_01 | NCMGGAWGYN | ttttTCAGa | 1.244487 | -872 | 0.85 | 0.86 |
| | LM02COM_01 | SNNCAGGTGNNN | agacaTCTGgga | 1.242813 | -857 | 0.82 | 0.89 |
| | MYOD_Q6 | NCANCTGNY | gaCATCtgag | -0.147777 | -856 | 0.92 | 0.89 |
| | SRY_02 | NWNAACAAWANN | actATGttttt | 3.860390 | -824 | 0.81 | 0.85 |
| | CDPCR3HD_01 | NATGATSSS | taATGATtct | 2.348950 | -822 | 0.89 | 0.93 |
| | OCT1_02 | NGAAATTATCANNNN | ttaaGGA Tttggg | 8.0359815 | -810 | 0.86 | 0.86 |
| | GATA_C | NGATAAGNMNN | gGATATggggc | 1.411097 | -805 | 0.87 | 0.86 |
| | USF_Q6 | GYCAGCTGNC | ggCACCtGat | 13.858419 | -793 | 0.82 | 0.89 |
| | USF_C | NCACGTGN | ccACCTtga | 0.507662 | -792 | 0.86 | 0.83 |
| | USF_Q6 | GYCAGCTGNC | gggtGtGtGgc | 6.857788 | -684 | 0.82 | 0.87 |
| | CETSTP54_01 | NCMGGAWGYN | gcAGGAGaag | 1.244487 | -676 | 0.93 | 0.88 |
| | USF_Q6 | GYCAGCTGNC | ggGACAGat | 6.857788 | -667 | 0.86 | 0.85 |
| | CETSTP54_01 | NCMGGAWGYN | acAGCAtaca | 1.244487 | -664 | 0.93 | 0.91 |
| | GATA_C | NGATAAGNMNN | gGATACaaga | 1.411097 | -661 | 0.88 | 0.89 |
| | CETSTP54_01 | NCMGGAWGYN | ccAGGAAatg | 1.244487 | -578 | 0.93 | 0.92 |
| | GATA_C | NGATAAGNMNN | agaATGccat | 1.411097 | -552 | 0.87 | 0.94 |
| | OCT1_06 | CWINAWTKWSATRYN | gATATggcaat | 8.438364 | -551 | 0.94 | 0.85 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité ducore | Similarité de la matrice |
|-------------|-------------------|------------------|--------------|----------|-------------------|-------------------|--------------------------|
| | LMO2COM_01 | SNNGAGGTGNNNN | atgCATGtgtcc | 1.242813 | -559 | 0.82 | 0.89 |
| USF_Q6 | | GYCACGTGNC | tgcatTGTGtc | 6.857788 | -538 | 0.86 | 0.86 |
| USF_C | | NCACGTGN | 9CATGTGt | 0.507862 | -537 | 0.88 | 0.93 |
| USF_C | | NCACGTGN | 9CATGTGt | 0.507862 | -537 | 0.82 | 0.85 |
| OCT1_06 | CWNATKWSATRYN | atttttttaaATTTta | 8.438384 | -512 | 0.89 | 0.90 | 0.90 |
| TATA_C | NCATAAAAR | aaTTTAAAAA | 8.111772 | -504 | 0.93 | 0.87 | 0.87 |
| MYCMAX_02 | NANCACGTGNNW | tatctGTGgtg | 3.484391 | -483 | 0.90 | 0.86 | 0.86 |
| USF_Q6 | GYCACGTGNC | ggCACACgecc | 6.857788 | -477 | 0.86 | 0.86 | 0.86 |
| CETS1P54_01 | NCMGGAWGYN | tctAGGAGga | 1.244487 | -453 | 0.93 | 0.91 | 0.91 |
| PADS_C | NGTGGTCTC | actGTATtc | 5.230232. | -404 | 0.90 | 0.91 | 0.91 |
| CETS1P54_01 | NCMGGAWGYN | gatttCCAGg | 1.244487 | -401 | 0.85 | 0.89 | 0.89 |
| CAAT_01 | NNNRRCCAATSA | gttagCCACtct | 3.434507 | -377 | 0.83 | 0.85 | 0.85 |
| NFY_Q6 | TRRCCAAATSRN | gagCCACtctc | 5.187369 | -375 | 0.81 | 0.85 | 0.85 |
| TATA_C | NCATAAAAR | tTTTTAAAAA | 16.345002 | -360 | 0.93 | 0.88 | 0.88 |
| TATA_C | NCATAAAAR | tTTTTAAAAa | 16.345002 | -360 | 0.93 | 0.88 | 0.88 |
| AP2_Q6 | MKCCCSCTNGGG | gtccgtGGGag | 12.284970 | -337 | 0.98 | 0.85 | 0.85 |
| CETS1P54_01 | NCMGGAWGYN | acAGGAatgt | 1.244487 | -320 | 0.93 | 0.85 | 0.85 |
| OCT1_06 | CWNATKWSATRYN | gCCATtttcaggatg | 8.438384 | -294 | 0.83 | 0.86 | 0.86 |
| CEBPB_01 | RNRTKNGMAAKNN | ccatTTCAaaatgt | 1.460836 | -293 | 0.99 | 0.92 | 0.92 |
| E47_01 | NSNGAGGTGKNCNN | citACAGtgcacccg | 9.748242 | -276 | 0.83 | 0.85 | 0.85 |
| USF_Q6 | GYCACGTGNC | cttCACAGtg | 6.857788 | -276 | 0.86 | 0.86 | 0.86 |
| USF_Q6 | GYCACGTGNC | ccatGtGtGac | 13.888419 | -274 | 0.82 | 0.89 | 0.89 |
| USF_C | NCACGTGN | acAGtGtGta | 0.507862 | -273 | 0.86 | 0.92 | 0.92 |
| ARP1_01 | TGACCCYTGTGANCYWW | tgacccGTtGcccc | 123.979856 | -268 | 0.83 | 0.87 | 0.87 |
| CETS1P54_01 | NCMGGAWGYN | gtAGGAAatcc | 1.244487 | -217 | 0.93 | 0.88 | 0.88 |
| CETS1P54_01 | NCMGGAWGYN | ggaaTCCtGt | 1.244487 | -214 | 0.93 | 0.88 | 0.88 |
| VMYB_01 | AYAAACGGNN | tctTGTGtgt | 3.427439 | -210 | 0.82 | 0.86 | 0.86 |
| CEBPB_01 | RNRTKNGMAAKNN | ccttaaGAAACcc | 1.460836 | -200 | 0.99 | 0.89 | 0.89 |
| USF_C | NCACGTGN | gtAGGTGtc | 0.507862 | -130 | 0.86 | 0.92 | 0.92 |
| GATA_C | NGATAAGNNNN | gtgtcaATATCc | 1.411097 | -126 | 0.87 | 0.89 | 0.89 |
| OCT1_02 | NNGAATATKCANNN | tgcaatATCCaaatag | 8.039815 | -125 | 0.86 | 0.92 | 0.92 |
| NFKB_C | NGGGACTTCCA | ggggaaatATGCC | 42.843021 | -43 | 0.93 | 0.86 | 0.86 |
| NFKAPPAB_01 | GGGAAATTYCC | gaaatccccc | 10.361187 | 45 | 0.90 | 0.86 | 0.86 |
| CETS1P54_01 | NCMGGAWGYN | ccTGGAGatc | 1.244487 | 74 | 0.85 | 0.85 | 0.85 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similité du core | Similité de la matrice |
|-------------|---------------------|---------------------|-------------------|----------|-------------------|------------------|------------------------|
| | RFX1_01 | NNGTNRCCNNRGYAAACNN | acgtTTCCTggatattt | 7.172878 | 117 | 0.88 | 0.85 |
| | CETSIIP54_01 | NCMGGAWGYN | gggtTCCAGG | 1.244487 | 135 | 0.85 | 0.86 |
| USF_C | NCACGTGN | ccAGGTGc | 0.507662 | 140 | 0.86 | 0.92 | 0.92 |
| CEBPB_01 | RNRTKNNGMAAKNN | gagggtGGAAaaa | 1.460836 | 181 | 0.87 | 0.88 | 0.88 |
| LMO2COM_01 | SNNCAGGTGNNN | ggatCTCggic | 1.242873 | 216 | 0.82 | 0.88 | 0.88 |
| MYOD_Q6 | NNCANCNTGNY | gtCATCtggt | -0.147777 | 217 | 0.92 | 0.93 | 0.93 |
| CETS1P54_01 | NCMGGAWGYN | cacatTCAGt | 1.244487 | 253 | 0.85 | 0.91 | 0.91 |
| USF_Q6 | GYCACGTGNC | cttcAGTGic | 6.857788 | 256 | 0.86 | 0.86 | 0.86 |
| CEBPB_01 | RNRTKNNGMAAKNN | cactTTCaaaatga | 1.460836 | 267 | 0.99 | 0.93 | 0.93 |
| SRF_Q6 | GNCCAWATAWGGMN | ttCCAAatggggcc | 30.107806 | 271 | 0.87 | 0.91 | 0.91 |
| GC_01 | NRGGGGGGGGGCGNK | accccccACCCccc | 35.805371 | 289 | 0.87 | 0.91 | 0.91 |
| SP1_Q6 | NGGGGGGGGGGYN | ccccccACCCccc | 25.529462 | 290 | 0.82 | 0.92 | 0.92 |
| CCT1_06 | CWNAWTKWSATRYN | caAAATttacilia | 8.438364 | 309 | 0.89 | 0.85 | 0.85 |
| CEBPB_01 | RNRTKNNGMAAKNN | acitatacAAAGaa | 1.460836 | 317 | 0.99 | 0.94 | 0.94 |
| OCT1_05 | MKNATTGCTATYY | ccaaatttCATgtia | 49.364942 | 363 | 0.85 | 0.86 | 0.86 |
| CEBPB_01 | RNRTKNNGMAAKNN | cagtTTCCttatcg | 1.460836 | 405 | 0.99 | 0.87 | 0.87 |
| GATA_C | NGATAAGNNNN | tttccCTATCg | 1.411097 | 408 | 0.89 | 0.93 | 0.93 |
| GATA_C | NGATAAGNNNN | tatcgGTATCa | 1.411097 | 414 | 0.88 | 0.89 | 0.89 |
| USF_Q6 | GYCACGTGNC | ggCACAGcaca | 6.857788 | 433 | 0.86 | 0.85 | 0.85 |
| AP2_Q6 | MKCCCSCNNGCCG | ttccggGGGcaa | 12.284970 | 448 | 0.98 | 0.85 | 0.85 |
| OCT1_06 | CWNAWTKWSATRYN | gAAAATaaaataacg | 8.438364 | 457 | 0.89 | 0.86 | 0.86 |
| CEBPB_01 | RNRTKNNGMAAKNN | aaataacGAAAtca | 1.460836 | 463 | 0.99 | 0.91 | 0.91 |
| AP2_Q6 | MKCCCSCNNGCCG | c当地AGGGcg | 12.284970 | 546 | 0.86 | 0.91 | 0.91 |
| RFX1_02 | NNGTNRCCNNNRGYAACNN | tggtTTCCttggggat | 7.174515 | 574 | 0.88 | 0.89 | 0.89 |
| RFX1_01 | NNGTNRCCNNRGYACNN | tggtTTCCttggggat | 7.172878 | 574 | 0.88 | 0.88 | 0.88 |
| LMO2COM_01 | SNCAGGTGNNN | ggccAGCtgcata | 1.242873 | 640 | 0.88 | 0.94 | 0.94 |
| LMO2COM_01 | SNCAGGTGNNN | ggccaaGCTGca | 1.242873 | 640 | 0.88 | 0.91 | 0.91 |
| MYOD_Q6 | NNCANCNTGNY | gcacGCTGca | 1.709898 | 641 | 0.92 | 0.97 | 0.97 |
| MYOD_Q6 | NNCANCNTGNY | gcCAGGCGca | 1.709898 | 641 | 0.92 | 0.90 | 0.90 |
| AP1_C | NTGASTCAN | atGACACAG | 1.430304 | 652 | 0.86 | 0.85 | 0.85 |
| USF_Q6 | GYCACGTGNC | c当地ACgggg | 6.857788 | 670 | 0.82 | 0.85 | 0.85 |
| USF_Q6 | GYCACGTGNC | gcACCCgggg | 13.858419 | 680 | 0.82 | 0.88 | 0.88 |
| USF_C | NCACGTGN | cCACCTGgg | 0.507662 | 681 | 0.86 | 0.92 | 0.92 |
| RFX1_02 | NNGTNRCCNNRGYAAACNN | ctggggggatGGAAAccg | 7.174515 | 685 | 0.88 | 0.89 | 0.89 |

| Filtration | Site | Consensus | Séquence | score Z | Position/TSS (pb) | Similarité du core | Similarité de la matrice |
|-------------|-----------------|-----------------|--------------|-----------|-------------------|--------------------|--------------------------|
| | AP2_Q6 | MKCCCCSCNGGGCG | agCCCCGaggcc | 12.284970 | 734 | 0.98 | 0.85 |
| OCT1_Q6 | NNNNATGCAAATNAN | aaaaatgcAAACagg | 11.430842 | 781 | 0.80 | 0.88 | |
| HNF3B_01 | NNNTRTTTTRYTY | atgcAAACAGgg | 2.929849 | 785 | 0.99 | 0.92 | |
| CETS1P54_01 | NCMGGAWGYN | gctTCAGa | 1.244487 | 796 | 0.85 | 0.89 | |
| CDPCR3HD_01 | NATYGATSSS | ttCATCata | 2.349950 | 814 | 0.93 | 0.93 | |
| CDPCR3HD_01 | NATYGATSSS | catcGATAgc | 2.349950 | 816 | 0.84 | 0.95 | |
| GATA_C | NGATAAGNMNN | cGATAAGccct | 1.411097 | 819 | 0.89 | 0.88 | |
| CETS1P54_01 | NCMGGAWGYN | ccatTCAGc | 1.244487 | 825 | 0.85 | 0.89 | |
| HNF3B_01 | NNNTRTTTTRYTY | actTGTtgcgg | 2.929849 | 907 | 0.99 | 0.88 | |
| CETS1P54_01 | NCMGGAWGYN | tttgTCCTgt | 1.244487 | 1027 | 0.93 | 0.86 | |

Tableau 3

| Promoter | Site | Transcription Factor | Core Similarity | Matrix Similarity | Z score | Position/TSS (bp) | motif | consensus |
|----------------|------|----------------------|-----------------|-------------------|---------|-------------------|--------------------------|------------------------|
| ABCA7 Human | A | GFI-01 | 1.00 | 0.88 | 4.43 | -389 | GGCACTATAATCGGAGACTCTAGA | NNNNNNAAATCANGNNNNNNNN |
| | B | HNF3B_Q3 | 0.99 | 0.85 | 4.37 | -547 | GAATGTGGGCC | NNNTRTRTRY |
| | C | CEBP_Q1 | 0.97 | 0.85 | 2.13 | -498 | CGTTGTGAAATGA | RNRTRKNGMAAKNN |
| | D | CEBP_Q1 | 0.87 | 0.85 | 2.13 | -469 | ATCTAGTGGAAAC | RNRTRKNGMAAKNN |
| | E | NF1_Q6 | 1.00 | 0.86 | 2.00 | -402 | GCCTGGCCAGCCCCCGGG | NNNTGGGNNNNNNCCNNNN |
| | F | AP4_Q5 | 1.00 | 0.90 | 1.68 | -340 | TGGAGCGGT | NNCAGCTGNN |
| | G | NFKAPPAB_Q1 | 1.00 | 0.80 | 9.96 | -280 | GGGAACGTGCG | GGGAMITYCC |
| | H | NFY_Q6 | 1.00 | 0.89 | 2.00 | -108 | CGCCCAATAGC | TRRCGAAATSRN |
| ABCA7 Mouse | A | GFI-01 | 1.00 | 0.88 | 2.96 | -342 | TTGGCTACAACTCAGGCAACTATT | NNNNNNAAATCANGNNNNNNNN |
| | B | HNF3B_Q3 | 0.89 | 0.85 | 3.25 | -825 | AACATATTGATTC | NNNTRTRTRY |
| | C | CEBP_Q1 | 0.99 | 0.84 | 1.72 | -787 | TGATTCGAAATG | RNRTRKNGMAAKNN |
| | D | CEBP_Q1 | 1.00 | 0.91 | 1.72 | -750 | ATGTTGCTAAATG | RNRTRKNGMAAKNN |
| | E | NF1_Q6 | 1.00 | 0.88 | 2.61 | -688 | TCTGGCTGGTGGCAGGA | NNNTGGGNNNNNNCCNNNN |
| | F | AP4_Q5 | 1.00 | 0.93 | 2.03 | -386 | CACACCACTG | NNCAGCTGNN |
| | G | NFKAPPAB_Q1 | 1.00 | 0.88 | 11.12 | -301 | GGAGGCTGCG | GGGAMITYCC |
| | H | NFY_Q6 | 1.00 | 0.89 | 5.63 | -156 | CCTCGAAATGGC | TRRCGAAATSRN |

Tableau 4

Oligonucleotides spécifiques du gène ABC-A7 humain.

| Nom | Séquence (5'-3') | Orientation |
|----------|--------------------------|-------------|
| ABCA7_U2 | CTTCAGCCCGACCGTTG | Sense |
| ABCA7_AJ | AGAATTTCATGTATCGCC | Sense |
| ABCA7_L2 | CGATGGCAGTGGCTTGTGTTGG | Antisense |
| ABCA7_L1 | GCGGAAAGCAGGTGTTGTTCAC | Antisense |
| ABCA7_AL | CTGGAGTTGCTGTCAGAG | Sense |
| ABCA7_AK | GGGTAAAAGGTGTATCTGG | Antisense |
| ABCA7_AN | TCACGGAGGACCAATAAGATC | Sense |
| ABCA7_AM | TGTCAGTGTACGGAGTAG | Antisense |
| ABCA7_AP | CCTGGAAGCTGTGTGC | Sense |
| ABCA7_AO | ACGGAGACGCCAGGAC | Antisense |
| ABCA7_AR | GTCCTGGCGTCTCCGTTTC | Sense |
| ABCA7_AQ | CTCGTCCAGGATAACAAAC | Antisense |
| ABCA7_AT | GTGCTGCCCTACACGG | Sense |
| ABCA7_AS | CAGTGCCCAGCCCTGTAC | Antisense |
| ABCA7_AV | ACCCCAGAGTCTCCATCC | Sense |
| ABCA7_AU | GAGAAGCCTCCGTATCTGAC | Antisense |
| ABCA7_AX | CTGCTCTCCCTGCTGTTGC | Sense |
| ABCA7_AW | GCACCATGTCAATGAGCC | Antisense |
| ABCA7_AZ | CCTCAGCATGGGATACTG | Sense |
| ABCA7_AY | GCTTGCCTTGTCTCCCTC | Antisense |
| ABCA7_BA | ACCCACGGCTTCTCTCC | Antisense |
| ABCA7_Q | AGCCAGCAACGCAATCCTCC | Sense |
| ABCA7_R | CGCACCATGTCAATGAGCCC | Antisense |
| ABCA7_L3 | TGAAGACGTGCGGTGCG | Antisense |
| ABCA7_L4 | TGTCTCCGGCGATACTGAAATTG | Antisense |
| ABCA7_L5 | ACCTCAGACCCAGACCCCTTACGC | Antisense |
| ABCA7_U4 | GGAATGAGGTCAGAAAGGG | Sense |
| ABCA7_U5 | ATGCAAGTTCCCTGGGAGTTAG | Sense |
| ABCA7_U6 | CTCCTTCCGGTGAATGTTGACG | Sense |
| | | |

REVENDICATIONS

1. Acide nucléique comprenant un polynucléotide ayant au moins 20
5 nucléotides consécutifs de la séquence nucléotidique choisie parmi les séquences
SEQ ID N°1-5, ou un acide nucléique de séquence complémentaire.

2. Acide nucléique ayant au moins 80% d'identité en nucléotides avec un
acide nucléique selon la revendication 1.

10 3. Acide nucléique hybride, dans des conditions d'hybridation de forte
stringence, avec un acide nucléique selon la revendication 1 ou 2.

15 4. Acide nucléique selon l'une des revendications 1 à 3, capable de
moduler la transcription d'un polynucléotide placé sous son contrôle.

20 5. Acide nucléique selon la revendication 4, comprenant un polynucléotide
allant du nucléotide en position -1 au nucléotide en position -1111 par rapport au
premier nucléotide transcrit, localisé en position 1112 de la séquence nucléotidique
SEQ ID N°1.

6. Acide nucléique selon la revendication 4, capable d'activer la
transcription d'un polynucléotide d'intérêt placé sous son contrôle.

25 7. Acide nucléique selon la revendication 4, capable d'inhiber la
transcription d'un polynucléotide d'intérêt placé sous son contrôle.

8. Acide nucléique comprenant:
a) un acide nucléique selon l'une des revendications 1 à 7; et
30 b) un polynucléotide codant pour un polypeptide ou un acide nucléique
d'intérêt.

9. Acide nucléique selon la revendication 8, caractérisé en ce que l'acide
nucléique d'intérêt est un oligonucléotide de type sens ou antisens.

10. Vecteur de clonage et/ou d'expression recombinant comprenant un acide nucléique selon l'une des revendications 1 à 9.

5 11. Cellule hôte transformée par un acide nucléique selon l'une des revendications 1 à 9 ou par un vecteur recombinant selon la revendication 10.

10 12. Mammifère transgénique non humain dont les cellules somatiques et/ou les cellules germinales ont été transformées par un acide nucléique selon l'une des revendications 1 à 9 ou par un vecteur recombinant selon la revendication 10.

13. Procédé pour le criblage d'une substance ou d'une molécule modulant la transcription du polynucléotide constitutif de l'acide nucléique selon la revendication 8, caractérisé en ce qu'il comporte les étapes suivantes:

15 a) mettre en culture une cellule hôte transformée selon la revendication 11;
b) incuber la cellule hôte transformée en présence de la substance ou molécule candidate;
c) détecter l'expression du polynucléotide d'intérêt;
d) comparer les résultats de détection obtenus à l'étape c) avec les
20 résultats de détection obtenus par mise en culture de la cellule hôte transformée en l'absence de la molécule ou substance candidate.

14. Kit ou nécessaire pour le criblage *in vitro* d'une molécule ou substance candidate modulant la transcription du polypeptide d'intérêt codé par un
25 polynucléotide constitutif de l'acide nucléique selon la revendication 8, comprenant :
a) une cellule hôte transformée selon la revendication 11;
b) le cas échéant, les moyens nécessaires à la détection de la transcription
du polynucléotide d'intérêt constitutif de l'acide nucléique selon la revendication 8.

30 15. Procédé de criblage *in vivo* d'une substance ou molécule modulant la transcription d'un polynucléotide d'intérêt constitutif de l'acide nucléique selon la revendication 8, caractérisé en ce qu'il comprend les étapes suivantes:
a) administrer la substance ou molécule candidate à un mammifère transgénique non humain selon la revendication 12;

b) détecter l'expression du polynucléotide d'intérêt chez le mammifère transgénique tel que traité à l'étape a);

c) comparer les résultats de détection de l'étape b) aux résultats observés chez un mammifère transgénique non humain selon la revendication 12 n'ayant pas 5 reçu l'administration de la substance ou molécule candidate.

16. Kit ou nécessaire pour le criblage *in vivo* d'une molécule ou substance candidate modulant la transcription du polynucléotide d'intérêt constitutif de l'acide nucléique selon la revendication 8, comprenant:

10 a) un mammifère transgénique non humain selon la revendication 12;

b) le cas échéant, les moyens nécessaires à la détection de la transcription dudit polynucléotide d'intérêt.

17. Substance ou molécule modulant la transcription d'un polynucléotide 15 d'intérêt constitutif de l'acide nucléique selon la revendication 8.

18. Substance ou molécule selon la revendication 17, caractérisée en ce qu'elle est sélectionnée selon le procédé de la revendication 13 ou de la revendication 15.

20

19. Composition pharmaceutique comprenant, en tant que principe actif, une substance ou une molécule selon l'une des revendications 17 ou 18.

25

20. Composition pharmaceutique selon la revendication 19, caractérisée en ce qu'elle est destinée au traitement et/ou à la prévention des déficiences dans le métabolisme des lipides, ou dans les mécanismes impliquant le système immunitaire et l'inflammation.

30

21. Substance ou molécule selon l'une des revendications 17 ou 18, en tant

que principe actif d'un médicament.

22. Procédé de détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant les étapes suivantes:

- a) extraire l'ARN messager total à partir d'un matériel biologique provenant du sujet à tester;
- b) quantifier l'ARN messager de ABCA7 présent dans ledit matériel biologique;
- 5 c) comparer la quantité d'ARN messager de ABCA7 obtenue à l'étape b) avec la quantité d'ARN messager de ABCA7 attendue chez un sujet normal.

23. Procédé de détection d'une altération de la transcription du gène ABCA7 chez un sujet, comprenant les étapes suivantes:

- 10 a) séquencer, à partir d'un matériel biologique provenant du sujet à tester, un polynucléotide localisé en amont du site d'initiation de la transcription du gène ABCA7;
- b) aligner la séquence nucléotidique obtenue en a) avec la séquence SEQ ID N°1;
- 15 c) déterminer les différences nucléotidiques entre le polynucléotide séquencé provenant du matériel biologique du sujet à tester et la séquence SEQ ID N°1 de référence.

24. Kit ou nécessaire pour la détection d'une altération de la transcription 20 du gène ABCA7 chez un sujet, comprenant les moyens nécessaires à quantifier l'ARN messager de ABCA7 dans un matériel biologique provenant dudit sujet à tester.

25. Kit ou nécessaire pour la détection d'une altération de la transcription 25 du gène ABCA7 chez un sujet, comprenant les moyens nécessaires au séquençage d'un polynucléotide localisé en amont du site d'initiation de la transcription du gène ABCA7 chez le sujet à tester.

26. Procédé de criblage d'une molécule ou substance modulant la 30 transcription du polynucléotide d'intérêt constitutif de l'acide nucléique selon la revendication 8, comprenant les étapes suivantes:

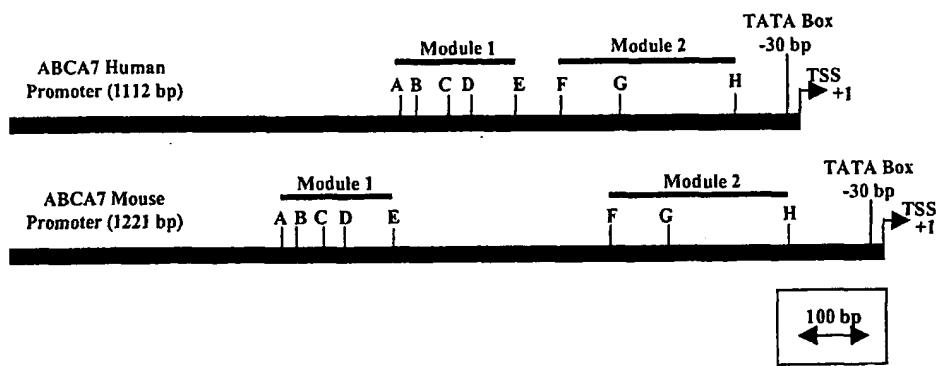
- a) incuber un acide nucléique selon l'une des revendications 1 à 9 ou un vecteur recombinant selon la revendication 10 avec une molécule ou substance candidate à tester;

b) détecter le complexe formé entre la molécule ou substance candidate et la molécule ou la substance candidate.

27. Kit ou nécessaire pour le criblage d'une molécule ou substance candidate modulant la transcription du polynucléotide d'intérêt constitutif de l'acide nucléique selon la revendication 8 comprenant :

- 5 a) un acide nucléique selon l'une des revendications 1 à 9 ou un vecteur recombinant selon la revendication 10;
- 10 b) le cas échéant, les moyens nécessaires à la détection du complexe formé entre la molécule ou la substance candidate et ledit acide nucléique.

Figure 1



- A: GFI-01 (GFI1)
- B: HNF3B-03
- C: CEBP-01
- D: CEBP-01
- E: NF1-Q6
- F: AP4-Q5
- G: NFκB-01
- H: NFY-Q6

Figure 2**SEQ ID NO: 1**

-911 AAAACCTCTGTTGTACGAAGAGAAGGTGCCAAGAGAGTTGGCGTCGATGAGGGCGTGC
 -851 TTTGCTTGATGCTTGTGGGAGAGAGGGAGGTCTGGGGATGGGGATCAAGGGGA
 -791 AAATGTCCACCTCACCATGGGAGGAGGAGCAAAGCTGAAGCCACAGGTGAGTCTGGT
 -731 GGAATGAATGATTGAAGGCCGGGACTTGGGTAGAGGGAGAGGGCTGGCTTCCCTGGC
 -671 ATTTGGAGAAGAGGCAGTTCCCTCAAATGCCCTCATGCGCTTGGCTGCACCTACCTT
 -611 ACAGCGCAAGTCTCGTGGCCTAGCCTGGATGTCTCCCGTTGGCGAACCTCTATTATC
 -751 CTCAAAGCCCCAACGGAATGCCACCTCCCTGCCGCGGGAGCCGTCACGCCCTCACT
 -691 CTCCCCAGCGCCTCAAAGCTGTGGACCCACACGCTCCATTTCAGCTCACCTCCAGCC
 -631 TGAAGAGTTATTCAACTCTTCCAGAGTGGAAACGGTTTCCTCAAAATCAGGG
 -571 GFI1 HNF3B
 TAGCCACTATAATCGGAGACTCTAGAATGTTGGCCCCCTCCCCCTCTGCCATCCTCTGC
 -511 CEBPB CEBPB
 AGAACCGAGGAGCGTGTGGAATGAATGAACGAATGATCTAGTGGAACCCCTAC
 -451 NF1
 TTTACAGACGGACGAGTGTAGTCCCAGAGTCTGGACTAAACTAGAGGGAGCCTGGCCAGC
 -391 AP4
 CCCGGGGACAGCGGGACAGAGGAACTCCTGCAATTGGAGCTGGTATTGCAGCCGG
 -331 TTATACAACGTGGGAGGCAGCCTGGCTCCCAAAGACAGCGCAGCCTCGTCCGGAGG
 -271 NFKAPPAB
 GCGGCCTGCCCTGGACCTGCCGGCACTCCGCCACCCCTACGGTATGCAGCAAGAGCCGC
 -211 GCGGTCCCTTAAGAAACCCGGCTAGGCAGGGCCTTCTGTGATCCGTCTCCCTT
 -151 NFY
 GCCCGCGCAGCTCCGACGGAGCAGGCCAGTGAGTGACGGCAGGTGCCCCAATAGCAGCG
 -91 TGCAGAGGCAGGGCGTGCCCCGGCGCTGCTACCTGCGCGGGCAAGCTAGCGCACTTGG
 -31 TATA +1
 CTTAAGGGCGCGCTCCCTGCCTGCTGGAGGGAAAGGCGGAGACTCGAGAATGCG
 +30
 GAGCCCCCTGGAAGGTGAGAAGGACTCGGAGAGGGAAAGAAGGCCGAGACTCGAGAATGCG
 +90
 GGGTTGGGGCCGGAGGGATGCAAGTTCCCTGGGAATTAGGGGTCCAGCCTCTGACCTC
 +150
 CTTCCGGTGAATGTTGACGACGGCTGAATTGATCACTGATTCTCAAGGGGGCATCGGAC
 +210
 ATCTGGGACCCCTTAAGAGGGCCTTGCCGATCACACACCTGCAGCCCCCTGCCGTTAGA
 +270
 ACTCCTGCACTCCCCCTGCCCGTCTTACAAATGGAGAAACTGAGCCCACCTCCCCCAGA

Figure 2 (suite)

+330
TCCTAAGTCCCCTTGATGTAAAGGAAAGAACCTGGCGTAAGGGTCTGGGTCTGAGGTC
+390
CCAGTTCCGGCCTGGTCACCTTAGCAACTTCCTGCCCTCTGTCAGCGTCAGATTCTCC
+450
ATCTGTGTCAAGAGGTGGACCGGCCAAGGAAAATAGATCAGGAATCGCTGACTCCAGGAG
+510
TCTCTATCCCAGCCCCCTCGCCTGACTCTTCTCTGGCTCCCGCGGTCCCTCTGAGCGAT
+570
TAATGCTACATAAGGTGTGGGCAGAGCTGGGTCGTGCCTCCAGCTGGCAACTGCCTGT
+630
CTCTCTGGGTGCCTGGGTTGCTTCCTTGGCCTCGGTTCCACTTCTGTAGAGTGGGTT
+690
GATAGTCCAGCAGTCCCTGGCGTGTGAAATGTCCAGCAGTGCCTGGGCAACTGGGCTG
+750
TTATCTCGGAGAACAGTGAGGGAAAGGAATCCTTGCCTGGGCTGGGCAAGGAGGAGG
+810
CTGGGGGTCAAGGACCTGGAAGAGGCTTCCAGGTGAGGCTGGGTTGGAGCCTGGTGACGA
+870
AAGCGTTAACGCCAAACTCGGTCCCTGGAGGATTAGAGGATGATCTTAAGTCCCCAGCT
+930
GTCAGCCCTGCTCAGAGCGACAGTCCTGGCAGCCAATCAGATGCGAGGACGGCTGCGGGT
+990
TGCGCTCCATTGGTTACTCCACCCCTGGGTAGCGGAGCCTTTATCGAGTGA
+1050
TGTGCGCTCGCTTAATCAGAGCTCCAGGAACCCCTGCGCTGTGGATAAAGGAATGAG
+1110
GTTCAAGAAAGGGGCAGGGAGTTGCCCGCAGCCGACCGCACGTCTCAGCCGACCGTTG
+1170
TCCTGACCTCTGTCCCCTGCCAGTCTCACCATGG

Figure 3

SEQ ID NO: 4

-1220 AAAACAAAAAAACAAAACAAAACAAAACAAAACATAAAAACCTCTGTTCTA
 -1160 AGAGTAAAGTACATTCCCTGAGTTGGCGTGATGGAGGGGGCGCTGTCTAGAAGCAAGGT
 -1100 GCAAGCCCTGCACAAAGTTAGGGAGAAGGCGAGAACAGACACAGTTGAATGAATGATG
 -1040 TGAGATAGATCGGGCTAGGGTGGAGAAAGAGGCTGAGTCTCCCTACCAGCTTCCTCG
 -980 AACTCCTATGCATCTGCAAAACCCCAACTCTAAGGCCCTAAGTCACCGCTGCCAGGG
 -920 TGATCTACACCCATCTCCCTATGCTTGAAATAAACAGTTTTTTCCAGAGT
 -860 GF11 HNF3B
 AGGAGACATCTGAGAATCTTGCTACAATCCAGGCAACTATTGATTCTAATCTTAGGATA
 -800 CEBP CEBP
 TTGGGCTGCCACCTGATTCTGAAATTGTCTAGACCAGAGGATGTTGCTAAATGAATGTG
 -740 NF1
 CAGGTCTTGAAGCTCTACTTGGAGATGAGCTCACAGAGGCTGTGGTACAATTCTGGCT
 -680
 GGTGGCAGGAGATGGCACAGGATACAAAGACCTTGTGCAAACCTTCCGACCTAAACTTGG
 -620 TCTTGCCTGAGGTCCCACATCATGGTAGGCAAGAATAGACTCCAGGAAATGGCCTCTG
 -560
 ACCTCCACAGATATGCCATGCATGCATGTGTCCTACCCCTAATAAGCAAATTAAATT
 -500 TAAAAACAAAGTTACTTGTGGTGGCACACGCCCTTAATCCAGCACTCAGGAGGCAGAG
 -440 AP4
 GCAGGGCGGATCTCTGTGAGACCAGCCTGGCTAAGCAGTGATTCCAGGCCTACCACAGC
 -380
 AGTGTGAGCCACTCTCAAAATTAAAAAGTATTTTAAAAAGGAGTCCTGGGAGAGGAG
 -320 NFKappaB
 ACAGGAATGTCTTGCTGTGGGAGCTGCCATTCAAGATGTGAACTCACAGGTGACCCGT
 -260
 TGTCCCCCTTTGTCGTGTCCTCAGTGAAGCCAAACTGATGCAGCAGGAATCCTGTTGTC
 -200 NFY
 CCTTTAAGAAACCCGGCTCGGAGAGGCGGGCTGTGGTCCCGCCTCCAAATGGCAAAGT
 -140
 CGCCTGAGTAGCAGGTGCAATATCCAATAGTAGCGTTAGGGGGCGGGCTGGGTGCTCCT
 -80 TATA ^{bos}
 TAGGGCACCGGGTTGCGAAGGGCGTCGTCCGAATTGAGCGGGCTCCACTTAAAGGGGC
 -20 +1
 CGCGCTCCCCCGCCGAGGCCGAGAGGAGCGAAAGTGGATGGAGTTGGGGCCTCAGAAC
 +40
 GGTGAGAAAATCCCCGAGAGGGTGGAAAGGTGGAGCCCTGGAGATCTGGGATGCTGTGGG
 +100
 TGAGGGTGGCTGAGCCACGTTCCCTGATAATTGGGGTCCAGGTGCCTACTCTCCCTT
 +160
 GCCCTTCCTTATCCTCCGGGGAGTGTGGAAAAATGGACCACCGATCCTCACAGCGGTC

Figure 3 (suite)

+220 ATCTGGTCACCTCGAGGGACCTCTGCCAACCTACACCTCCAGTGTCCCACTTCCAAATG
+280 AGGCCTGTCACCCCCCACCCCCCAGATCTAAATTCACTTTATGAAAGAAAAAGTCCC
+340 CGAGTGGAAAGCCGCCAATTCCATGTAGATGGTTAAACTTGGCAATTCCCTCTGTC
+400 AGCCTCAGTTCCCTATCGGTATCATGAAGCAGGCCACAGGCATAACAGTTCCGGGGAA
+460 ATAAAATAACGAAATCAGGAATGGCGTGCTCAAGGAGCCTGTCCCTGACTCCTCCTAGCC
+520 GGCAGGTCTTCTGTACCCCTCTTGACTCCGGAGGGCGGGCCCTCCTTCTCTGGTTT
+580 CCTTGGGAGCGTGACTTGGCCCTTTGAGCCTCAGTTCCATCTCTTAAAAATAGAA
+640 GGCAGCTGCAAATGACACAGACTCCGGTCTCACCGGGGCCACCTGGGGCAACGGAA
+700 CCGAGACCCCGGTCTGGTATGAGGCTGACCGTGGAGCCCCGAGCCCCAAGCCCCAAGCTT
+760 TAAACCCAAGCTCCGCCCTAAGAATGCAAACAGGGTCTTCCAGACCCAGCCTTCATC
+820 GATAGCCCTTCCAGCCAATCAGCTACGAGGACGGCTGCGCGCCGGTTCCATTGGTCAC
+880 TTCCCTAGTGAATTCTTCTATGGTGCCTTGTGCGGGCTCTTGCGGGAGATTAT
+940 TGAGGCTCAGCCGATGTTCGGAAGGATGAGGATCAGAGACCCGAGACATTGTCTGGA
+1000 GCCACACAGCTCACTCTCAGCCTTTCTTGTCCCTGCCTCCGCTGTTCCGGTCCAG
+1060 AGTCATCATGG

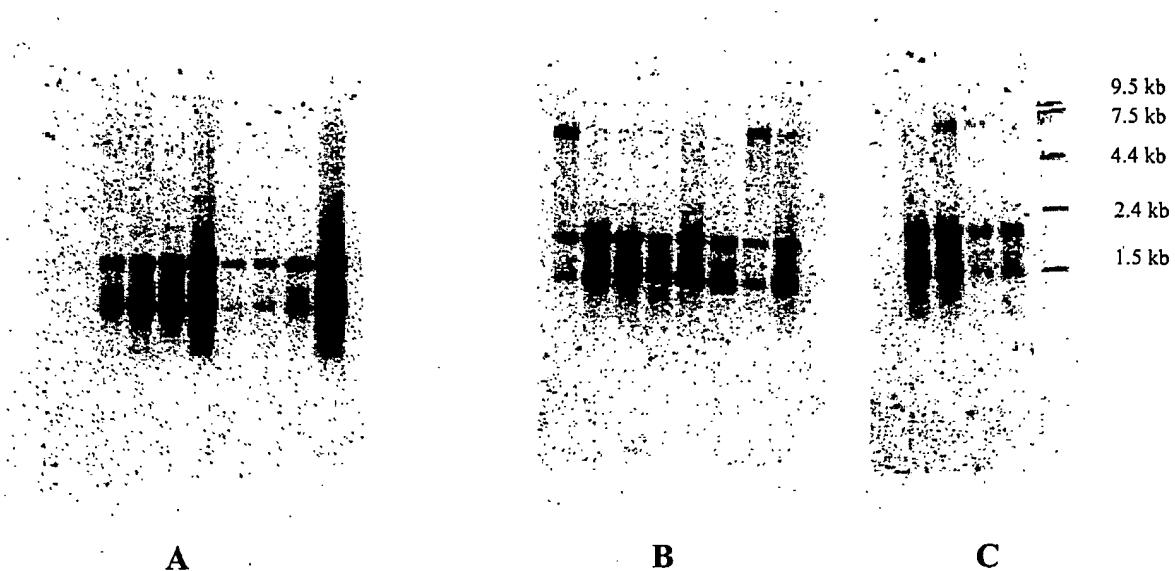
6/7

Figure 4

1 2 3 4 5 6 7 8

1 2 3 4 5 6 7 8

1 2 3 4

**A****B****C**

1. Pancreas
2. Rein
3. Muscle
4. Foie
5. Poumon
6. Placenta
7. Cerveau
8. Cœur

1. Leucocyte
2. Colon
3. Petit intestin
4. Ovaire
5. Testicule
6. Prostate
7. Thymus
8. Rate

1. Rein fœtal
2. Foie fœtal
3. Poumon fœtal
4. Cerveau fœtal

7/7

Figure 5

B H Li Ln Lu S K A T U

ABC-A7



B: Cerveau
H: Cœur
Li: Foie
Ln: Ganglions
Lu: Poumon
S: Rate
K: Rein
A: Glandes Surrénales
T: Thymus
U: Utérus



Figure 6

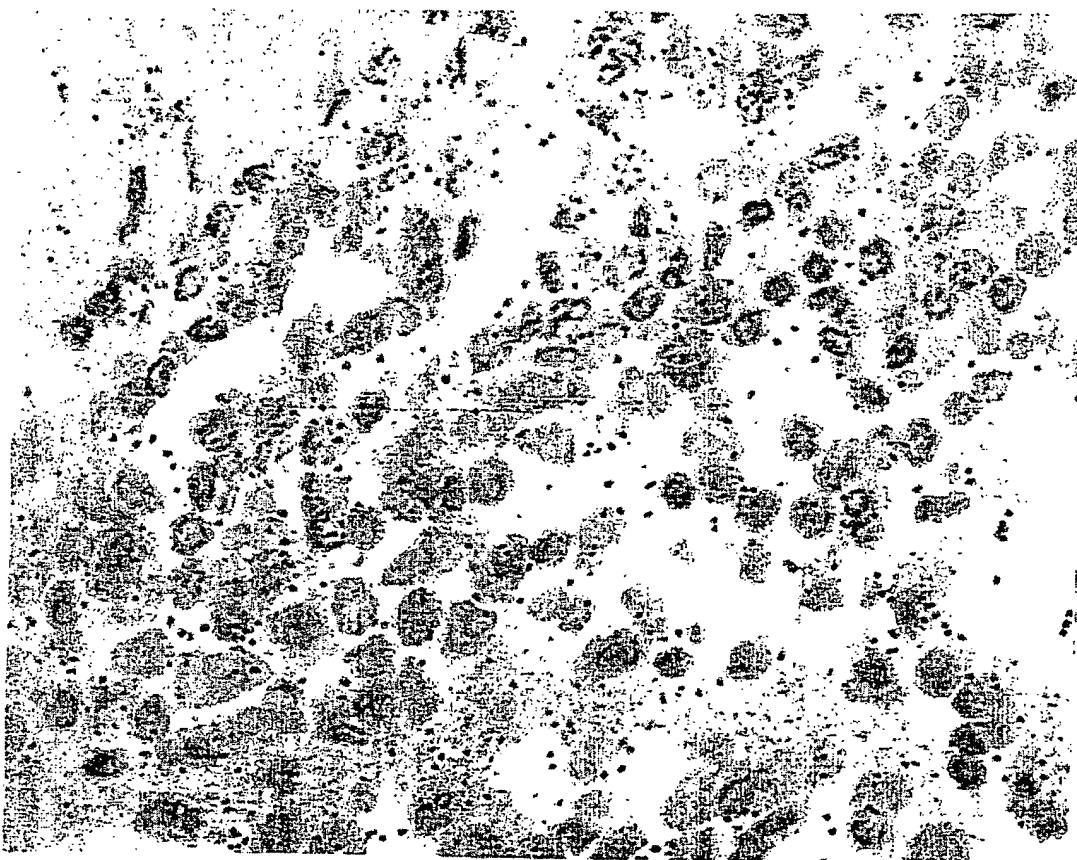


Figure 7

10/13

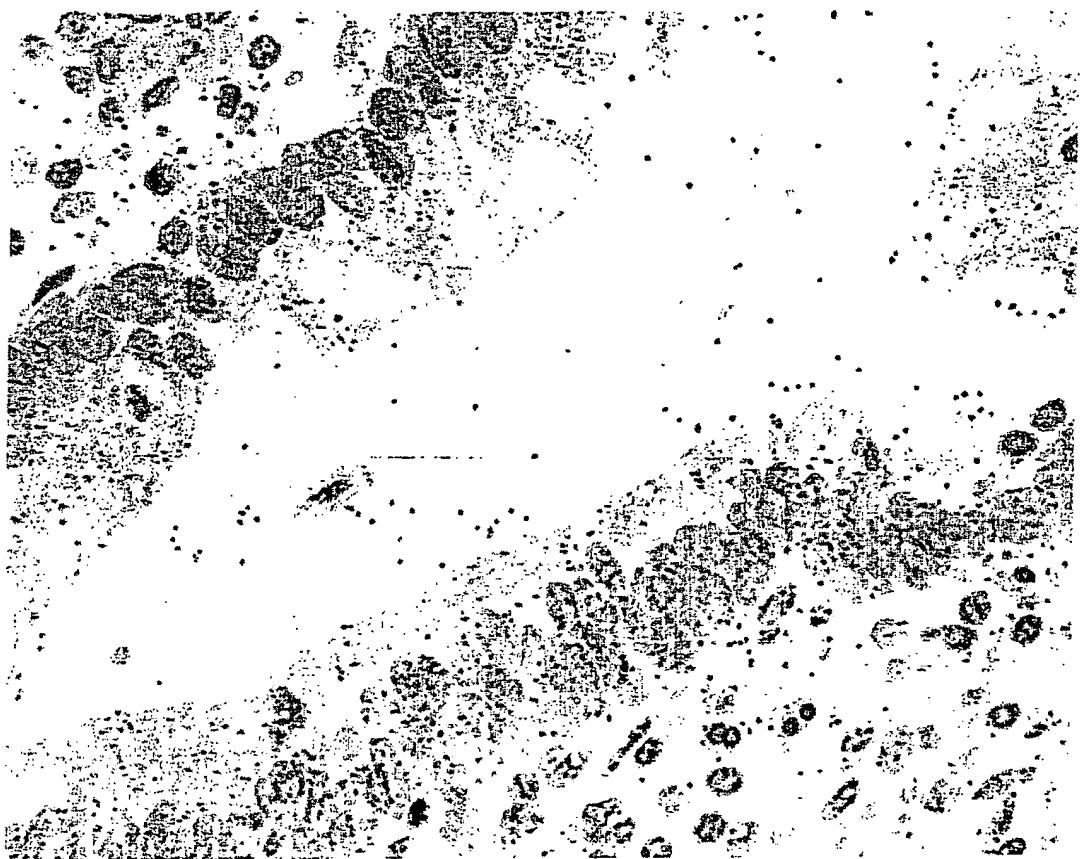


Figure 8

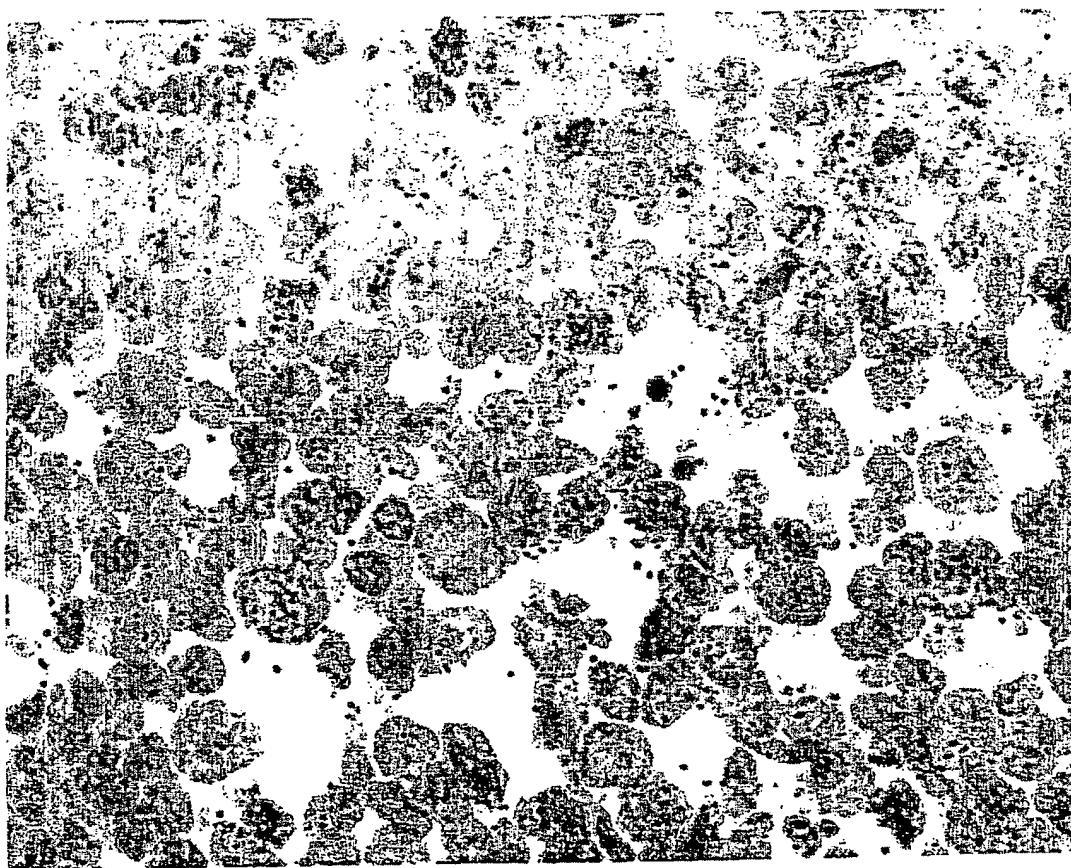


Figure 9

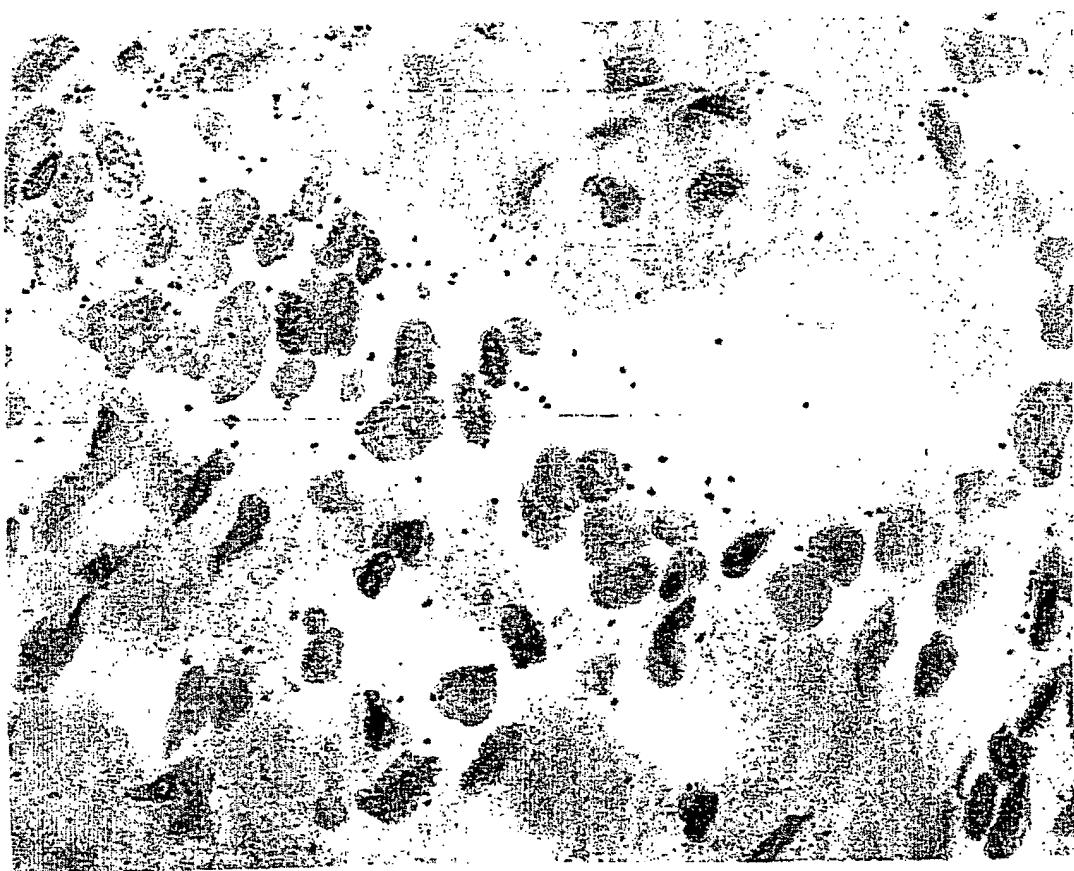


Figure 10

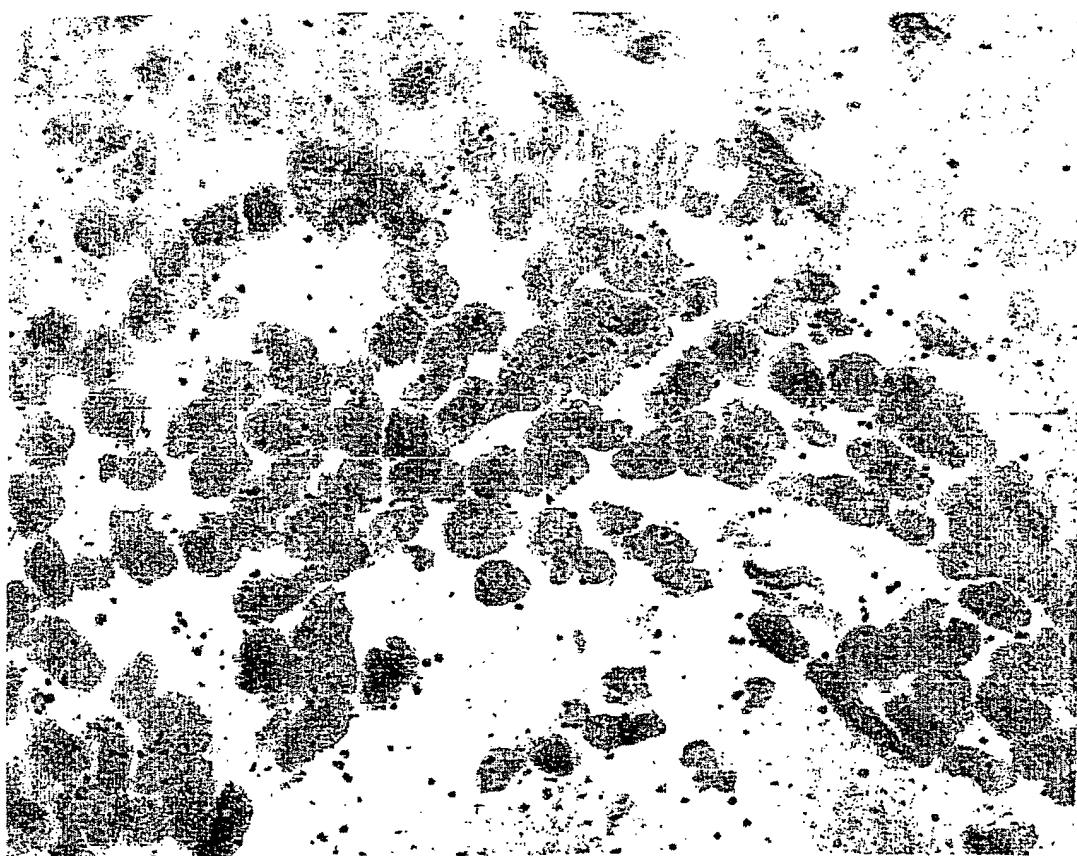


Figure 11

LISTE DE SEQUENCES

<110> AVENTIS PHARMA SA , INSERM

<120> ACIDE NUCLEIQUE REGULATEUR DU GENE ABCA7, MOLECULES
MODULANT SON ACTIVITE ET APPLICATIONS THERAPEUTIQUES

<130> promoteur ABCA7

<140>

<141>

<160> 12

<170> PatentIn Ver. 2.1

<210> 1

<211> 2322

<212> ADN

<213> Homo sapiens

<400> 1

aaaacccctctg ttgtaccaa gagaagggtgg ccaagagagt tggcgctcgat gagggcgtgc 60
tttgctttga tgctttgtg gggagagagg aggtcttggg ggatgggggg atcaagggga 120
aaatgtccac ctcaccattg ggaggaggag caaaagctga agccacagggt gagtctgggt 180
ggaatgaatg atttgaaggg cccggacttg gggtagaggg agaggctggg cttcctggcc 240
atttggagaa gaggcagttc cctcaaatgc ccccatgcg ctttggctgc actctacatt 300
acagcgcaag tctcggtggcc tcagcctgga tgtctccccg ttggcgaact cctatttatac 360
ctcaaagccc caacggcaat gccacccctt gcccggggag ccgtccccac gcctctact 420
ctccccagcg cttccaaagc tgtggaccca cacgtccca tttcagcttc acctccagcc 480
tgaagagttt atttcaactc ttcttccaga gtggaaacg gtttttcctc aaaatcaggg 540
tagccactat aatcgagac tctagaatgt tggcccccctc cccctctcgatcact 600
agaagccgag gagcgttcggt ggaatgaatg aatgaacgaa tgatctagtg gaaccctac 660
tttacagacg gacgagtgtt gtcccagagt ctggactaaa ctagagggag cctggccagc 720
ccgggggaca gccccggacag agggaaactcc tgcaattcgg, agctgcggta ttgcagccgg 780
ttataacaacg tggggaggca gcctggctcc ccaaagacag cgcagctcg ttccggagg 840
gcggcctgccc tgggacctgc cgggactcc gccacccctac ggtgatgcag caagagccgc 900
gcgggtccctt taagaaaccc ggcttagggca ggccttctcg tgatcccgtc tcctcccttg 960
gccccggcag ctccgacggaa gcaggccagt gagtgcggg caggtcgccc aatagcagcg 1020
tgcagaggca ggggcgtgcc cgggcgtgc tacgtgcgc gcaagctca ggcacttgg 1080
cttaaggggc ggcgcgtcc ctgcctgtcg ctggggcggag ggaaggcggc aagagctgcg 1140
gagccctgg aaggtgagaa ggactcgaggg agggaaagaag gcccggact cgagaatgcg 1200
gggttggggc cgggagggtat gcaagttccc tggaaattag ggggtccagc ctctgaccc 1260
cttccgggtga atgttgcattt gatcaactgtat tctcaagggg ggcacccggac 1320
actcctgcac tcccccttgc cccgtcttac aaatggagaa actgagccca ctccccccaga 1380
tcctaagtcg cgttgcattt gatcaactgtat tctcaagggg ggcacccggac 1440
tcctaagtcg cgttgcattt gatcaactgtat tctcaagggg ggcacccggac 1500

ccagttccgg cctggtcacc tttagaact tcctgcctt ctgtcagcgt cagattctcc 1560
 atctgtgtca gaggtggacc ggcccaagga aaatagatca ggaatcgctg actccaggag 1620
 tctctatccc agcccccctcg cctgactctt tctctggctc ccgcggctcc tctgagcgtat 1680
 taatgtaca taaggtgtgg gcagagctgg ggtcgtgcct ccagctggc aactgcctgt 1740
 ctctctgggt gcctgggttt gccttcttgg gcctcggttt ccacttctgt agagtgggt 1800
 gatagtcacg cacttcccct gggcgtgtga aatgtccacg actgccaata ttgcgtgctg 1860
 ttatcttcgg agaacagtga gggaaagga atccctgcct gggctggcc aggcaggagg 1920
 ctgggggtca ggacctggaa gaggcttcca ggtgaggctt ggggtggagc ctgggtgacga 1980
 aagcgttaag cccaaactcg gtccctggag gattagagga tgatcttaa gtcccagct 2040
 gtcagccctg ctcagagcga cagtcctggc agccaatcag atgcgaggac ggctgcccgt 2100
 tgcgccttcca ttggtttact ccacccctgg ggtagcggag cctctttatc gagtgactac 2160
 tgttgcctc gctctaatac gagttccag gaaccctgcg ctgtggata aaggaatgag 2220
 gttcagaaag gggcaggagg ttgcccgcag ccgcaccgcg cgtttcagc ccgaccgtt 2280
 tcctgacctc tctgtcccgt cccctgccccca gtctcaccat gg 2322

<210> 2

<211> 1111

<212> ADN

<213> Homo sapiens

<400> 2

aaaaaccttg ttgtacgaa gagaaggctgg ccaagagagt tggcgtcgat gaggcggtgc 60
 ttgtcttga tgctttgtg gggagagagg aggtcttggg ggatgggggg atcaagggga 120
 aatgtccac ctcaccattt ggaggaggag caaaagctga agccacaggt gagtctgggt 180
 ggaatgaatg atttgaaggg ccgggacttg ggttagaggg agaggctggg ctgcgtggcc 240
 atttggagaa gaggcagttc cctcaaattgc ccccatgcg ctggctgc actctacatt 300
 acagcgaag tctcggtggcc tcagcctgga tgtctccccg ttggcgaact cctatttatac 360
 ctcaaagccc caacggcaat gccacccctt gcccggggag ccgtccccac gcctctcact 420
 ctccccagcg cttcaaaagc tggcgtggcc cacgtccccat tttcagcttc acctccagcc 480
 tgaagagttt atttcaactc ttcttccaga gtggaaacg gttttccctc aaaatcaggg 540
 tagccactat aatcgagac tctagaatgt tggcccccctc cccctctgc catcctctgc 600
 agaagccgag gagcgttcgt ggaatgaatg aatgaacgaa tgatcttagt gaaaccctac 660
 ttacagacg gacgagtgtt gtcccagagt ctggactaaa cttagagggag cctggccagc 720
 cccggggaca gccccggacag agggaaactcc tgcaattcgg agctgcggta ttgcagccgg 780
 ttataacaacg tggggaggca gcctggctcc cccaaagacag cgcagcctcg ttcccccggagg 840
 gccccctgccc tgggacctgc cgggcactcc gccaccctac ggtgatgcag caagagccgc 900
 gccccctttaaagaaaccc ggcttaggcga ggccttctg tgatcccgtc tcctcccttgc 960
 gccccggcagc ctccgacggc gcaggccagt gagtgcggg caggtcgccc aatagcagcg 1020
 tgcagaggca gggcgtgccc ccggcgtgc tacctgcgcg ggcaagctca ggcacttgg 1080
 cttaaggggc ggcgcgtcc ctgcctgtc c 1111

<210> 3

<211> 1211

<212> ADN

<213> Homo sapiens

<400> 3

tgggcggagg gaaggcggca agagctgcgg agccctgga aggtgagaag gactcgaga 60
 gggagaagg cccgagactc gagaatgcgg gttggggcc gggaggatg caagttccct 120
 gggatttagg ggtccagcc tctgaccc ttcgggtgaa ttttgacgac ggctgaattt 180
 atcactgatt ctcaaggggg gcatcgaca tctggaccc ttaagagggc ctttgcgtat 240
 cacacacctg cagcccttg cccgttagaa ctctgcact ccccttgcc cctgtttaca 300
 aatggagaaa ctgagccac tcccccagat cctaagtccc gttgtatgtt aaggaagaa 360
 ccctggcgta agggctggg tctgagggtcc cagttccggc ctggtcaccc tttagcaactt 420
 cctgccttc tgcagcgtc agattctcca tctgtgtcag aggtggaccg gccaaggaa 480
 aatagatcag gaatcgctga ctccaggagt ctctatccca gccccttcgc ctgactctt 540
 ctctggctcc cgcggccct ctgagcgatt aatgtacat aaggtgtggg cagagctggg 600
 gtcgtgcctc cagctgggca actgcctgtc tctctgggtg cctgggtttt ctttctggg 660
 cctcggttcc cacttctgtt gagtgggtt atagttccagc acttcccttg ggcgtgtgaa 720
 atgtccagca ctgccaatat tcgttgctgt tatcttcgga gaacagttag gggaaaggaa 780
 tccttgccctg ggctgggcca ggcaggaggc tggggtcag gacctggaaag aggcttccag 840
 gtgaggcttgggagcc tggtgacgaa agcgtaagc ccaaactcgg tccctggagg 900
 attagaggat gatcttaag tcccccagctg tcagccctgc tcagagcgac agtccctggca 960
 gccaatcaga tgcgaggacg gctgcgggtt ggcgttccat tgggttactc caccctggg 1020
 gtagcggagc ctctttatcg agtactact gtttgactcgt ctctaatcag agcttccagg 1080
 aaccctgcgc tggggataa aggaatgagg ttcaaaaaagg ggcaggaggat gcccgcagc 1140
 cgcaccgcac gtctcagcc cgcaccgttgt cctgacccctc ctgtccctgc ccctgcccag 1200
 tctcaccatg g 1211

<210> 4

<211> 2291

<212> ADN

<213> Mus musculus

<400> 4

aaaaacaaaaaa aaaaaacaaa aacaaaacaa aaacaaaaac aataaaaaacc tctgtttcta 60
 agagtaagt acattcctga gtttggccgt gatggagggg ggcgtgtcta gaagcaaggt 120
 gcaaggccctg cacaaggat agggagaagg cgagaaacag acacagttga atgaatgtt 180
 tgagatagat cggggctagg gtggagaaag aggctgagtc tccctcacca gcttcctcg 240
 aactcctatg catctgaaa accccaactt ctaaggcccc ctaactcagc cttgcccagg 300
 tgcacatcac ccatctccct ctagctttt gaaataaaac cagttttttt tttccagagt 360
 aggagacatc tgagaatctt gcctacaatc caggcaacta ttgattctaa tcttaggata 420
 ttgggctgcc acctgattct gaaattgtct agaccagagg atgttctaa aatgaatgtt 480
 caggtccttg aagctctact ttggagatga gtcacagag gctgtgttac aattctggct 540
 ggtggcagga gatggcacag gataaaaaaa ctttgtcaa accttccgac ctaaaacttgg 600
 tcttccttg aggtcccaca tcatgttggg caagaataga ctccaggaaa tggccctctg 660
 acctccacag atatgcacatg catgcacatg tccatcccta ataagcaaat taattaaatt 720
 taaaaacaaa gtttacttgtt ggtggcacac gcctttaatc ccagcactca ggaggcagag 780
 gtcaggccat ctctgtgaga ccagccctgtt ctaagcactg atttccaggc ctaccacagc 840
 agtgtgagcc actctcaaaa taaaaaaatgta tttttttttt ggagtccttg gggagaggag 900
 acaggaatgtt cttgtgtgg ggagctgcca tttcaagatg tgaactcaca ggtgaccctgt 960

tgtccccctc tttgtcggtgt cccagtgaaag ccaaactgat gcagcaggaa tcctgttgta 1020
 cctttaagaac acccggtcg gagaggcggg ctgtggtccc gcctcctcca atggcaaagt 1080
 cgcctgagta gcaggtgcaa tatccaatag tagcgtagg gggcggggct gggtgctcct 1140
 tagggcaccg ggttgcgaag ggcgtcgta gcaattgagc ggggctccac ttaaaggggc 1200
 cgcgtcccc cgcgcaggcc gagaggagcg aaagtggatg gagtttgggg gcctcagaac 1260
 ggtgagaaaa tccccgagag ggtggaaagggt ggagcctgga gatctggggta tgctgtgggg 1320
 tgagggtggg ctgagccacg ttccctgata atttggggtt ccaggtgcct actctccctt 1380
 gccttcctt atccctccgg ggagtgtggg aaaaatggac cacgatcct cacagcggtc 1440
 atctggtcac ctcgaggac ctctgccaac ctacacctcc agtgtccac tttccaaatg 1500
 aggccgtgtca cccccccaccc cccagatctc aaatttcaact ttatgaaaga aaaaagtccc 1560
 cgagtggaaag cgcgcattt ccatgttagat ggttaaactt tggcaatttc cctctctgtc 1620
 agcctcagtt tccctatcggt tatcatgaaag cagccacag gcatacagtt cggggggaa 1680
 ataaaataac gaaatcagga atggcgtgct caaggagcct gtccctgact ctccttagcc 1740
 ggccgtcttc tgcgttcc ttttgactcc ggagggcggg cccttccttct tctctggttt 1800
 ccttgggagc gtgactttgc cccttttga gcctcagttc ccatactctta aaaaatagaa 1860
 ggccagctgc aaatgacaca gactccgggt ctcaccgggg gccacactggg gcgaacggaa 1920
 ccgagacccc ggtctggat gaggctgacc gtggagcccc gagccccaaag ccccaagcgtt 1980
 taaaaccaag ctccgcccccc taagaatgca aacagggtct tccagacccc agccttcac 2040
 gatagccctt ccagccaatc agctacgagg acggctgcgc gcccgggttcc cattggtcac 2100
 ttcccttagtg aatttcttct tatggtgctt tgttgcccgg gctctttgcg ggagatttat 2160
 tgaggcttag cccgatgttc ggaaggatga ggatcagaga cccgcagaca tttgtctgga 2220
 gcccacacagc tcactctcag cctttctt gtccctgtcct ctccgctgtt tccggccag 2280
 agtcatcatg g 2291

<210> 5

<211> 1220

<212> ADN

<213> Mus musculus

<400> 5

aaaacaaaaa aaaaaacaaa aacaaaacaa aaacaaaaac aataaaaaacc tctgtttctta 60
 agagtaaagt acattcctga gtttggccgt gatggagggg gcgctgtcta gaagcaaggt 120
 gcaaggccctg cacaacaaatggt agggagaagg cgagaaacag acacagttga atgaatgatg 180
 tgagatagat cggggcttagg gtggagaaag aggctgagtc tccctcacca gtttccttcg 240
 aactcctatg catctgcaaa accccaactt ctaaggcccc ctaactcactc cttgccaggg 300
 tgcatacact ccatctccct ctatgctttg tgaataaaac cagttttttt tttccagagt 360
 aggagacatc tgagaatctt gcctacaatc caggcaacta ttgattctaa tcttaggata 420
 ttgggctgcc acctgattct gaaattgtct agaccagagg atgttgctaa aatgaatgtg 480
 caggtccttg aagctctact ttggagatga gctcacagag gctgtggtac aattctggct 540
 ggtggcagga gatggcacag gataaaaaa gcttgcacaa accttccgac ctaaacttgg 600
 tctttgcctg aggtccccaca tcatggtagg caagaataga ctccaggaaa tggtcctctg 660
 acctccacag atatgccatg catgcacgtg tcctacccta ataagcaaat taattaaatt 720
 taaaaacaaa ggttacttgtt ggtggcacaac gcctttaatc ccagcactca ggaggcagag 780
 gcagggcggat ctctgtgaga ccagcctggt ctaagcactg atttccagggc ctaccacagc 840
 agtgtgagcc actctcaaaa ttaaaaagta tttttaaaaa ggagtccctt gggagaggag 900
 acaggaatgt ctgctgtgg ggagctgcca tttcaagatg tgaactcaca ggtgaccctgt 960

tgccccctc tttgtcggt cccagtgaag ccaaactgat gcagcaggaa tcctgttgc 1020
 ccttaagaa acceggctcg gagaggcggg ctgtggtccc gcctctcca atggcaaagt 1080
 cgcctgagta gcaggtgcaa tatccaatag tagcgtagg gggcggggct gggtgctcct 1140
 tagggcaccg ggttgcgaag ggctcgcc gcaattgagc ggggctccac taaaagggc 1200
 cgcgtcccc cgcggaggcc 1220

<210> 6
 <211> 1273
 <212> ADN
 <213> Homo sapiens

<400> 6
 tggcgagg gaaggcggca agagctcgag agccctgga aggtgagaag gactcgaga 60
 gggagaagg cccagactc gagaatcgag gggtgggccc gggagggatg caagttccct 120
 gggaaattagg gggtccagcc tctgacctcc ttccggtaa tggacgac ggctgaattt 180
 atcactgatt ctcaaggggg gcatcgacata tctggacccc ttaagaggcc ctttgcgc 240
 cacacacctg cagccccctg cccgttagaa ctccgtcaact ccccttgc cccgttttaca 300
 aatggagaaa ctgagccac tccccagat cctaagtccc gcttgatgt aaggaaagaa 360
 ccctggcgtt agggtctggg tctgagggtcc cagttccggc ctggtcaccc ttagcaactt 420
 cctgcccctc tgcagcgctc agattctcca tctgtgtcag aggtggaccc gcccaaggaa 480
 aatagatcg gaatcgctga ctccaggagt ctctatccca gcccccttgc ctgactcttt 540
 ctctggctcc cgccgtccct ctgagcgatt aatgtacat aaggtgtggg cagagctggg 600
 gtctggctc cagctggcact ctgcctgtc tctctgggtg cctgggttttgc ttttcttggg 660
 cctcggttttcc cacttctgtt gagtgggttgc atagccagc acttccccctg ggcgtgtgaa 720
 atgtccagca ctgccaatat tctttgttat tatcttgcga gaacagttag gggaaaggaa 780
 tccttgctc ggctggccca ggcaggaggc tgggggtcag gacctggaaagg aggcttccag 840
 gtgaggcttgggagcc tgggtacgaa agcgttaagc ccaaactcggttccctggagg 900
 attagaggat gatcttaag tccccagctg tcagccctgc tcagagcgac agtccctggca 960
 gccaatcaga tgcgaggacg gctgcgggtt ggcgtcccat tggtttactc caccctggg 1020
 gtagcggagc ctctttatcg agtgactact gtttgcctcg ctctaatcg agttccagg 1080
 aaccctgcgc tggggataa aggaatgagg ttcagaaagg ggcaggaggat tggccgcac 1140
 cgcaccgcac gtcttcagcc cgaccgttgtt cctgacccctc ctgtccctgc ccctgcccag 1200
 tctcaccatg gccttctgga cacagctgtat gctgtgtc tggaaaggatt tcatgtatcg 1260
 ccggagacag ccg 1273

<210> 7
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 7
 Met Ala Phe Trp Thr Gln Leu Met Leu Leu Leu Trp Lys Asn Phe Met
 1 5 10 15

Tyr Arg Arg Arg Gln Pro

<210> 8
<211> 7795
<212> ADN
<213> Homo sapiens

<400> 8
tgggcggagg gaaggcggca agagctgccc agccctggaa aggtgagaag gactcggaga 60
ggaaagaagg cccgagactc gagaatgcgg ggttggggcc gggaggatg caagttccct 120
ggaaattagg gggccagcc tctgacccctt ttccggtaa tgttgacgac ggctgaattg 180
atcaactgatt ctcaaggggg gcatcgacca tctgggaccc ttaaggggc ctttgcgc 240
cacacaccc cagccccctg cccgttagaa ctccctgact ccccttgcc ccgtcttaca 300
aatggagaaa ctgagccac tccccagat cctaagtccc gcttgcgtta aaggaaagaa 360
ccctggcgta agggctctggg tctgagggtcc cagttccggc ctggtcaccc ttgcactt 420
cctggccctc tgtcagcgtc agattctcca tctgtgtcag aggtggaccc gccaaggaa 480
aatagatcag gaatcgctga ctccaggagt ctctatccca gccccttcgc ctgactctt 540
ctctggctcc cgccggccct ctgagcgtt aatgtacat aagggtgtggg cagagctggg 600
gtcgtgcctc cagctggca actgcctgtc tctctgggtg cctgggtttt ctttcttggg 660
cctcggttcc aacttctgtt gagtgggggtt atagtcctcacttccctg ggcgtgtgaa 720
atgtccagca ctgccaatat tctgtgtgt tatcttcggaa gaacagttag gggaaaggaa 780
tccttgccctg ggctgggcca ggcaggaggc tgggggtcag gacccgttgcagg 840
gtgaggcttgggttgggagcc tgggtacgaa aegcttaagc ccaaactcgg tccctggagg 900
attagaggat gatcttaag tccccagctg tcagccctgc tcagagcgttccacttggca 960
gccaatcaga tgcaggagc gctgcgggtt ggcgttccat tggtttactc caccctggg 1020
gtagcggagc ctctttatcg agtgcactt gttgcctcg ctctaatcag agttccagg 1080
aaccctgcgc tggggataa aggaatgagg ttccagaaagg ggcaggagtttgcgc 1140
cgccaccgcac gtcttcagcc cgaccgttgt cctgacccctt ctgtcccggtc ccctgcccag 1200
tctcaccatg gccttctggaa cacagctgtt gctgtgtc tggaaattat tcatgtatcg 1260
ccggagacag ccggccctgc tccctggtaa attgtgtgg cctctttcc tcttcttcat 1320
cctgggtgtt gttccactt cccacccggc cctggagcac catgaatgc acttcccaa 1380
caagccactg ccattggcgcc gcaccgtgcc ctggctccag ggtctcatct gtaatgtgaa 1440
caacacccctgc ttccgcagc tgacaccggg cgaggagccc gggcgctga gcaacttcaa 1500
cgactccctg gtctcccgcc tgcgtccgtt gtcgtggact gtgtggag gggccactg 1560
ccacaggacg ctggctggcc tagggaaactt gatcgccacg ctggaggctgc cacgcac 1620
ggcccgactt caaccaacca agcagtctcc actggaaacca cccatgtgg atgtcgccgaa 1680
gtgtgtgtac tcactgtgtc gcacggaaatc cctgggggtt gcaactggcc aagcccgagg 1740
gccttgcac agttgttgg agggccgtgg ggacccgtgg caggagctcc tggcgctg 1800
cagccctggtaa gagcttccggg cactgtgtc gagaccccgaa gggaccagcg gcccctggaa 1860
gtgtgtgtca gagggccctt gcaactgtgtc gggacccgtgg agcactgtgg gcccctccct 1920
caactgggtac gaggctgtgtc acctgtatggaa gctgggtgggg caggagccag aatccggccct 1980
gccagacagc agcctgagcc ccgcctgtcc ggagctgtt gggccctgg acagccaccc 2040
gtgtcccgcc ctgtcttggaa gacccgtggaa gcctgtgtc ctggggaaactg tactttgc 2100
accagataca ccttttaccc ggaagctcat gggccagggtg aaccggaccc ttgcaggagct 2160
caccctgtgtc agggatgtcc gggagggtgtg ggagatgtgtc ggaccccgga tcttcaccc 2220

catgaaacgac agttccaatg tggccatgct gcagcggcgc ctgcagatgc aggatgaagg 2280
aagaaggcag cccagacactg gaggccggga ccacatggag gccctgcgat ctttctgga 2340
ccctggggc ggtggctaca gctggcagga cgcacacgct gatgtggggc acctgggtgg 2400
cacgctgggc cgagtgcacgg agtgcctgta cttggacaag ctggaggcgg caccctcaga 2460
ggcagccctg gtgtcgccgg ccctgcaact gctcgccgaa catcgattct gggccggcgt 2520
cgtcttcttg ggacctgagg actcttcaga ccccacagag caccctaaacc cagacctggg 2580
ccccggccac gtgogcatca aaatccgcattt ggacattgac gtggcacgaa ggaccaataa 2640
gatcagggac aggttttggg accctggcc accctggcc accctgaccc acctgcgtca 2700
cgtgtggggc ggcttcgtgt acctgcaaga cctgggtggag cgtgcagccg tccgcgtgt 2760
cagcggcgc aaccctccggg cccgcctta cctgcagcag atgcctatc cgtctatgt 2820
ggacgcgtg ttctctgcgtg tgctgagccg gtgcgtcccg ctcttcgtca cgctggcctg 2880
gatctactcc gtgacactga cagtgaaggc cgtggcgtgg gagaaggaga cgcggctg 2940
ggacaccatg cgcgccttgg ggctcaggccg cgggtgcgc tggctaggct gtttcctcag 3000
ctgcctcggg cccttcgtc tcagcgcgc actgctggttt ctgggtctca agctgggaga 3060
catcctcccc tacagccacc cggcgctgtt cttctgttc ttggcagccct tccgcgggtggc 3120
cacgggtgacc cagagcttcc tgcgcgc cttttatcc cgcgcctacc tggctgcggc 3180
ctgcggcggc ctggcctact ttccttcata cctgcctac gtgcgtgtt tggcttggcg 3240
ggaccgcgtg ccccgccggc gcccgcgtgc cgcgagccgc ctgtcgcccg tggcctcgg 3300
cttcggctgc gagagcttgg ctctgttggaa ggagcaggccg gagggcgcgc agtggcacaa 3360
cgtggcacc cggcctacgg cagacgttcc cagctggcc caggtctctg gccttcgt 3420
gctggacgcg ggcgtctacg gcctcgccac ctggtacccgtt gaagctgtgt gcccaggcca 3480
gtacgggatc cctgaaccat ggaattttcc ttttcggagg agctactgtt gggacactcg 3540
gcccccaag agtccagccc ctggccac cccgcgtggac ccaaagggtgc tggtagaaga 3600
ggcacccccc ggcctgatgc ctggcgatc ctggcgatc ctggagaagc gctttcctgg 3660
aagccgcag ccagccctgc gggggctcag cctggacttc taccaggccc acatcaccgc 3720
cttcctggc cacaacgggg cccgcaagac caccacccgtt tccatcttgc tggcctt 3780
cccaccatg ggtggctctg ctttcatctt gggccacgc gtccgcctca gcatggcgc 3840
catccggccc cacctggggc tctgttcata gtacaacgtt ctgtttgaca tgctgaecgt 3900
ggacgagcac gtctggttt atggggcgctt gaagggtctt agtgcgcgtt tagtggggcc 3960
cgagcaggac cgtctgttc aggtatgtgg gctggctcc aagcagatgt tgcagactcg 4020
ccacccctctt ggtggatgc aacggaaatgt gtccgtggcc attgcctttt tggggggctc 4080
ccaagtgtt atctggacg agcctacggc tggcgatc cctgttccccc gcccgggtat 4140
ttgggagctg ctgtcaaat accgagaagg tcgcacgtt atcctcttcca cccaccaccc 4200
ggatgaggca gagctgttgg gagaccgtgtt ggcgtggatc gcaggtggcc gttgtgt 4260
ctgtggctcc ccaactttcc tgcgcgtca cctggctcc ggctactacc tgacgtgtt 4320
gaaggcccgc ctggccctga ccaccaatgttga gaagggttgc actgacatgg agggcgtgt 4380
ggacaccagg caggaaaaga agaatggcag ccagggcagc agatcggca ctcctcgt 4440
gctggccctg gtacagactt gggtgccgg ggcacggctt gttggaggagc tgccacacga 4500
gctgggtgtt gtgtcgccctt acacgggttc ccatgacggc agtgcgttcc cactttccg 4560
agagctagac acgcggctgg cggagctgtt gctactggc tacggatctt ccgcacaccag 4620
cctcgaggag atttcttgc aggtgggttggaa ggagttgtgtt gctggacacag atatggagga 4680
tggcagctgc gggcagcacc tatgcacagg cattgtggc ttagacgttcc ccttacggct 4740
caagatgccg ccacaggaga cagcgcttggaa gaacggggaa ccagctgggtt cagccccaga 4800
gactgaccag ggctctgggc cagacgcgtt gggccgggtt cagggctggg cactgaccccg 4860
ccagcagctc caggccctgc ttctcaagcg ctttctgtt gcccgcgc gcccgcgg 4920
cctgttcgtcc cagatcgttgc tgcctgcctt ctttggggc ctggccctcg tggcgttcc 4980
catcgatgttcc ctttcggggc actacccggc tctgcggctc agtcccacca tgcgttcc 5040

tcagggtgtcc ttcttcagtg aggacgcccc aggggaccct ggacgtgccc ggctgctcga 5100
ggcgctgctg caggaggcag gactggagga gccccagtg cagcatagct cccacaggtt 5160
ctcgccacca gaagttcctg ctgaagtggc caaggtcttgc gccagtgccgca actggacccc 5220
agagtctcca tccccagctt gccagtgtag ccggcccggt gccccggcc tgctgcccga 5280
ctgcccggct geagctggtg gtcccccctcc gccccaggca gtgaccggct ctgggaaagt 5340
gttcaagaac ctgadaggcc ggaacctgtc tgacttcctg gtcaagacct acccgccct 5400
ggtgcccccag ggcctgaaga ctaagaagtg ggtgaatgag gtcagatacg gaggcttc 5460
gctggggggc cgagacccag gcctgcctc gggcaagag ttggccgct cagtgagga 5520
gttgtggcg ctgctgagtc ccctgcctgg cggggccctc gaccgtgtcc tgaaaaacct 5580
cacagcttgg gtcacacagcc tggatgtca ggacagtctc aagatcttgt tcaacaacaa 5640
aggctggcac tccatggtgg ctgttgc当地 ccgagccagc aacgcaatcc tccgtgtca 5700
cctgc当地 cggccggccc gccacgccc当地 cagcatcacc acactcaacc acccccttggaa 5760
cctcaccaag gaggcagctgt ctgaggctgc actgatggcc tccctggtgg acgttctcgt 5820
ctccatctgt gtggcttttg ccatgtcctt tgccccggcc agcttcactc ttgttctcat 5880
tgaggagcga gtcaacccgag ccaagcacct gcagctcatg gggggcctgt cccccaccct 5940
ctactggctt ggcaacttcc tctggacat gtgtactac ttggtgccag catgcategt 6000
ggtgctcata .tttctggccct tccagcagag ggcataatgtg gccccctgc当地 acctgcctgc 6060
tctctgctg ttgctactac tggatggctg gtcatcaca ccgctcatgt acccagccctc 6120
cttctcttc tccgtgccc当地 gcacacccata tggatggctc acctgcataaa acctctttat 6180
tggcatcaat ggaaagcatgg ccacccctgt gcttggctc ttctctgtc agaagctgca 6240
ggaggtgagc cgatcttga aacaggctt cttatcttcc ccccaacttct gcttggccg 6300
ggggctcatt gacatggtgc ggaaccaggc catggctgtat gccttggagc gcttgggaga 6360
caggcagttc cagtcacccca tgcgtgggaa ggtggctggc aagaaccttct tggccatgg 6420
gatacagggg cccctttcc ttctcttc当地 actactgtc cagcaccgaa gccaacttct 6480
gccacagccc agggtgaggt ctctgc当地 cctgggagag gaggacgagg atgtagcccg 6540
tgaacgggag cgggtggtcc aaggagccac ccaggggat gtgttggtgc tgaggaactt 6600
gaccaaggta taccgtggc agaggatgcc agctgttgc当地 cgcttgc当地 tgggattcc 6660
ccctgttag tggggggc当地 tgctgggtgt gaatggagca gggaaagacgt ccacgttcc 6720
catggtagc当地 ggggacacat tggccagcag gggcgaggct gtgctggcag gacacagcgt 6780
ggccccc当地 cccagtgctg cgacccctc当地 catggatac tggccatggataat ccgatgccat 6840
ctttgagctg ctgacggcc当地 gcaagcacct ggagctgtt ggc当地 ggc当地 ggggtgtccc 6900
ggaggccc当地 gttgccc当地 cccagtgctg cggccctggc当地 cgtctggac tctcatggta 6960
cgcaagaccgg cctgc当地 cccaggc当地 agggaaacaaa cgcaagctgg cgacggccct 7020
ggcgctggtt ggggacccag ccgtgggtt tctggacgag ccgaccacag gcatggaccc 7080
cagcgc当地 cggccctt当地 ggaacagctt tttggccctg gtgc当地 gggaggccgat 7140
gatgctcacc tccccatagca tggaggagtg tgaagcgtc当地 tgctc当地 gccc当地 tagccatcat 7200
ggtaatggg cgggtccctg ctgctggcag ccccaacat ctcaaggc当地 gattc当地 gggc 7260
gggtcacaca ctgacccctc当地 gggtggcc当地 cgcaaggtcc cagccggcag cggccctcgt 7320
ggcgccccc当地 tccctgggtt cggagctg cggaggccat ggaggccccc tgctc当地 tccctgggtt 7380
gtgccccc当地 ggagggc当地 ggc当地 ctggccctt当地 ggagagctgg cggtgacccgg 7440
cgcaagacac ggc当地 ggaggagg actttccctg gagccagacg atgctggagg aggtattttt 7500
gtacttctcc aaggaccagg ggaaggacga ggacaccgaa gagcagaagg aggccaggat 7560
gggagtgagc ccccgccag ggctgc当地 ccccaaaacgc gtcagccctg tccctcgatga 7620
cccttagcact gccgagactg tgctctgagc tccctcccc tgc当地 gggccggc当地 cggggaggcc 7680
ctggaaatgg caagggcaag gtagagtgcc taggagccct ggactc当地 tggc当地 gaggg 7740
gctggccccc当地 tggagaaaat aaagagaagg ctggagagaa ggc当地 tgggtgg .tgaaa 7795

<210> 9
<211> 20
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle: AMORCE

<400> 9
agccagcaac gcaatcctcc

20

<210> 10
<211> 20
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle: AMORCE

<400> 10
cgcaccatgt caatgagccc

20

<210> 11
<211> 22
<212> ADN
<213> Séquence artificielle

<400> 11
gcggaaagca ggtgttgttc ac

22

<210> 12
<211> 21
<212> ADN
<213> Séquence artificielle

<400> 12
cgatggcagt ggcttgttg g

21